

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

```

Run on:      July 4, 2003, 15:08:57 ; Search time 1505.66 Seconds
              (without alignments)
              9664.474 Million cell updates/sec

Title:      US-19-083-653B-52 copy 1_500
Perfect score: 500
Sequence:    1 gatagtgaagaagctca.....actaaaggtctttcaagct 500

```

```

Scoring table:  IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched:  2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters:  4109280

```

```

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 10%
                  Listing first 45

```

Database : GenEmbl : ☆

1: gb_bao
2: gb_btg.*
3: gb_in.*
4: gb_ov.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_ro.*
12: gb_sy.*
13: gb_un.*
14: gb_un.*
15: gb_vt.*
16: gb_vt.*
17: gb_vt.*
18: em_fun.*
19: em_fun.*
20: em_fun.*
21: em_fun.*
22: em_fun.*
23: em_fun.*
24: em_fun.*
25: em_fun.*
26: em_fun.*
27: em_fun.*
28: em_fun.*
29: em_fun.*
30: em_fun.*
31: em_fun.*
32: em_fun.*
33: em_fun.*
34: em_fun.*
35: em_fun.*
36: em_fun.*
37: em_fun.*
38: em_fun.*
39: em_fun.*
40: em_fun.*
41: em_fun.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Query			DB	ID	Description
			Match	Length	%			
C 1	500	100.0	1587	6	AX164746	Sequence		
C 2	500	100.0	1734	6	AX164761	Sequence		
C 3	500	100.0	1828	6	AX164744	Sequence		
C 4	500	100.0	1828	9	AF335584	Homo sapi		
C 5	500	100.0	2253	9	AF336376	Homo sapi		
C 6	500	100.0	3718	6	AX365307	Sequence		
C 7	500	100.0	3736	6	AX365305	Sequence		
C 8	500	100.0	4070	9	BC030645	Homo sapi		
C 9	500	100.0	160345	2	AC067870	Homo sapi		
C 10	500	100.0	180648	9	AP003043	Homo sapi		
C 11	500	100.0	182403	2	AC024052	Homo sapi		
C 12	496.8	99.4	3710	9	AY027518	Homo sapi		
C 13	496.8	99.4	3729	9	AY027517	Homo sapi		
C 14	482.8	96.6	1882	6	AX044487	Sequence		
C 15	482.8	96.6	1882	6	AX207483	Sequence		
C 16	460	92.0	485	6	AX331460	Sequence		
C 17	133.4	26.7	1428	9	BC033832	Homo sapi		
C 18	108.6	21.7	2870	10	BC030896	Mus muscu		
C 19	74.2	14.8	1472	6	AX044538	Sequence		
C 20	74.2	14.8	1472	6	AX207485	Sequence		
C 21	71	14.2	161951	2	CNS07PFD	Oryza sat		
C 22	64.8	13.0	253305	3	PFWAL3P7	Plasmodiu		
C 23	64.2	12.8	97643	9	AL157783	Human DNA		
C 24	64.2	12.8	171606	2	AL159118	Homo sapi		
C 25	62.8	12.6	176639	9	AC016987	Homo sapi		
C 26	62.8	12.6	182430	2	AC104589	Homo sapi		
C 27	62.8	12.6	189791	2	AC011629	Homo sapi		
C 28	62.6	12.5	40752	9	AC087065	Homo sapi		
C 29	62.2	12.4	62707	2	AC091197	Homo sapi		
C 30	62.2	12.4	186836	9	AC079340	Homo sapi		
C 31	61.6	12.3	191119	9	AC009957	Homo sapi		
C 32	61.4	12.3	256	6	AX302693	Sequence		
C 33	60.6	12.1	148202	2	AL772396	Danio rer		
C 34	60.2	12.0	253305	3	PFWAL3P7	Plasmodiu		
C 35	59.4	11.9	128769	2	AL139261	Homo sapi		
C 36	59.2	11.8	171317	9	AC020941	Homo sapi		
C 37	58.8	11.8	175928	2	AC129750	Rattus nor		
C 38	58.8	11.8	192177	2	AC109118	Rattus nor		
C 39	58.2	11.6	128834	2	AC090192	Homo sapi		
C 40	58.2	11.6	188971	9	AC007707	Homo sapi		
C 41	58	11.6	151498	9	AC099331	Homo sapi		
C 42	57.4	11.5	162075	9	HS127D3	Homo sapi		
C 43	56.8	11.4	175103	9	AC109135	Homo sapi		
C 44	56.6	11.3	61306	9	AC007423	Homo sapi		
C 45	56.6	11.3	162401	2	AC024024	Homo sapi		

ALIGNMENTS

RESULT 1					
LOCUS	AX164746/c				
DEFINITION	Sequence 3 from Patent WO0125437.	1587 bp	DNA	linear	PAT 22-JUN-2001
ACCESSION	AX164746				
VERSION	AX164746.1	GI:14545599			
KEYWORDS	human.				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 1587)				
AUTHORS	Shimkets,R.A., Lichenstein,H., Herrmann,J.L., Boldog,F.L., Minskoff,S. and Jeffers,M.				
TITLE	Growth factor polypeptides and nucleic acids encoding same				

JOURNAL Patent: WO 0125437-A 3 12-APR-2001;

Curagen Corporation (US)

FEATURES Location/Qualifiers

source 1..1587

/organism="Homo sapiens"

/db_xref="taxon:9606"

BASE COUNT 522 a 298 c 329 g 438 t

ORIGIN

Query Match 100.0%; Score 500; DB 6; Length 1587;

Best Local Similarity 100.0%; Pred. No. 6.3e-83;

Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTATATGTAAGAAAGCCTCATCTTTTGGATTTTAAATATACAAGATGCTTTCTTTAAGAGA 60

Db 1462 GTATATGTAAGAAAGCCTCATCTTTTGGATTTTAAATATACAAGATGCTTTCTTTAAGAGA 1403

Qy 61 GCAAGATTCAAAATGTTTGTGTTTCAAAATTTAAATAATTTATCTCCTAAATTTT 120

Db 1402 GCAAGATTCAAAATGTTTGTGTTTCAAAATTTAAATAATTTATCTCCTAAATTTT 1343

Qy 121 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAGAGATTTTAAGAGTCT 180

Db 1342 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAGAGATTTTAAGAGTCT 1283

Qy 181 AACTCAACATATGTAAGCTCTGCTGACCTGGTATATATATACCAAAAAAACAATTTTGTAT 240

Db 1282 AACTCAACATATGTAAGCTCTGCTGACCTGGTATATATATACCAAAAAAACAATTTTGTAT 1223

Qy 241 CTATATACATAGACATGAATATATTTCTGTGTGTTTGGCATATATAACCTCAAAAC 300

Db 1222 CTATATACATAGACATGAATATATTTCTGTGTGTTTGGCATATATAACCTCAAAAC 1163

Qy 301 ACTATTATTAATCAATCCCTATATTTCTAGGTATAGAGTTGATATACCTTTCTAC 360

Db 1162 ACTATTATTAATCAATCCCTATATTTCTAGGTATAGAGTTGATATACCTTTCTAC 1103

Qy 361 TTGCCATGGCATTAAACAAGCAAGCTGAGACTCAGCAACACCTTGTTGTCATTGCAATTG 420

Db 1102 TTGCCATGGCATTAAACAAGCAAGCTGAGACTCAGCAACACCTTGTTGTCATTGCAATTG 1043

Qy 421 CAGCTAGTAGTAGTTGGTTGCTGGTAGGAAAGGTCCTTATCTCACCTCCCTTAA 480

Db 1042 CAGCTAGTAGTAGTTGGTTGCTGGTAGGAAAGGTCCTTATCTCACCTCCCTTAA 983

Qy 481 ACTAAGCTCTTTTCAGGCT 500

Db 982 ACTAAGCTCTTTTCAGGCT 963

RESULT 2

AX164761/c

LOCUS AX164761 1734 bp DNA linear PAT 22-JUN-2001

DEFINITION Sequence 18 from Patent WO0125437.

ACCESSION AX164761

VERSION AX164761.1 GI:14545604

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1..1734

/organism="Homo sapiens"

/db_xref="taxon:9606"

BASE COUNT 530 a 365 c 382 g 457 t

ORIGIN

Query Match

100.0%; Score 500; DB 6; Length 1734;

Best Local Similarity

100.0%; Pred. No. 6.2e-83;

Matches 500; Conservative

0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTATATCTAAGAAAGCCTCATCTTTTGGATTTTAAATATACAAGATGCTTTCTTTAAGAGA 60

Db 1724 GTATATCTAAGAAAGCCTCATCTTTTGGATTTTAAATATACAAGATGCTTTCTTTAAGAGA 1665

Qy 61 GCAAGATTCAAAATGTTTGTGTTTCAAAATTTAAATAATTTATCTCCTAAATTTT 120

Db 1664 GCAAGATTCAAAATGTTTGTGTTTCAAAATTTAAATAATTTATCTCCTAAATTTT 1605

Qy 121 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAGAGATTTTAAGAGTCT 180

Db 1604 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAGAGATTTTAAGAGTCT 1545

Qy 181 AACTCAACATATGTAAGCTCTGCTGACCTGGTATATATATACCAAAAAAACAATTTTGTAT 240

Db 1544 AACTCAACATATGTAAGCTCTGCTGACCTGGTATATATATACCAAAAAAACAATTTTGTAT 1485

Qy 241 CTATATACATAGACATGAATATATTTCTGTGTGTTTGGCATATATAACCTCAAAAC 300

Db 1484 CTATATACATAGACATGAATATATTTCTGTGTGTTTGGCATATATAACCTCAAAAC 1425

Qy 301 ACTATTATTAATCAATCCCTATATTTCTAGGTATAGAGTTGATATACCTTTCTAC 360

Db 1424 ACTATTATTAATCAATCCCTATATTTCTAGGTATAGAGTTGATATACCTTTCTAC 1365

Qy 361 TTGCCATGGCATTAAACAAGCAAGCTGAGACTCAGCAACACCTTGTTGTCATTGCAATTG 420

Db 1364 TTGCCATGGCATTAAACAAGCAAGCTGAGACTCAGCAACACCTTGTTGTCATTGCAATTG 1305

Qy 421 CAGCTAGTAGTAGTTGGTTGCTGGTAGGAAAGGTCCTTATCTCACCTCCCTTAA 480

Db 1304 CAGCTAGTAGTAGTTGGTTGCTGGTAGGAAAGGTCCTTATCTCACCTCCCTTAA 1245

Qy 481 ACTAAGCTCTTTTCAGGCT 500

Db 1244 ACTAAGCTCTTTTCAGGCT 1225

RESULT 3

AX164744/c

LOCUS AX164744 1828 bp DNA linear PAT 22-JUN-2001

DEFINITION Sequence 1 from Patent WO0125437.

ACCESSION AX164744

VERSION AX164744.1 GI:14545598

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1..1828

/organism="Homo sapiens"

/db_xref="taxon:9606"

BASE COUNT 572 a 380 c 385 g 491 t

ORIGIN

Query Match

100.0%; Score 500; DB 6; Length 1828;

Best Local Similarity

100.0%; Pred. No. 6.1e-83;

Matches 500; Conservative

0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTATATCTAAGAAAGCCTCATCTTTTGGATTTTAAATATACAAGATGCTTTCTTTAAGAGA 60

Db 1818 GTATATCTAAGAAAGCCTCATCTTTTGGATTTTAAATATACAAGATGCTTTCTTTAAGAGA 1759

BASE COUNT	572 a	360 c	385 g	491 t
Query Match	100.0%	Score: 500;	DB 9;	Length 1828;
Best Local Similarity	100.0%;	Pred. No. 6.1e-83;		
Matches 500;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	GTATATGTAAGAAGCCTCATCTTTTGAATTTTAAATATACAAGATGCTTCTTTTAAGAGA	60	
Db	1818	GTATATGTAAGAAGCCTCATCTTTTGAATTTTAAATATACAAGATGCTTCTTTTAAGAGA	1759	
Qy	61	GCAAGATTCAAAATCTTTTGCTGTTTCAAAATTTAAATAAATATATCTCCCTAAATTTT	120	
Db	1758	GCAAGATTCAAAATGTTTGTGTTTCAAAATTTAAATAAATATATCTCCCTAAATTTT	1699	
Qy	121	CTAAGACATGTTTCATATATTTGACCATCCCTTATTTGGCAAGGATTTTAAAGAGTCT	180	
Db	1698	CTAAGACATGTTTCATATATTTGACCATCCCTTATTTGGCAAGGATTTTAAAGAGTCT	1639	
Qy	181	AACCAACAATATGTAAGCTCTGGTGTACCTGGTTATATATACCAAAAAACATTTGAT	240	
Db	1638	AACCAACAATATGTAAGCTCTGGTGTACCTGGTTATATATACCAAAAAACATTTGAT	1579	
Qy	241	CTATATACACATAGACATGAATATATTTCTGCTGTGTTTGTGCATATATAACCTCAAC	300	
Db	1578	CTATATACACATAGACATGAATATATTTCTGCTGTGTTTGTGCATATATAACCTCAAC	1519	
Qy	301	ACTATTATTAATGCAATCCTATATCTTGTAGGTATAGAAGTTGATGATATACCTTCTAC	360	
Db	1518	ACTATTATTAATGCAATCCTATATCTTGTAGGTATAGAAGTTGATGATATACCTTCTAC	1459	
Qy	361	TTGCCATGGCAATTAACAAAGCAAGGCTGAGACTCAGCAACCACTTGTGTTCATTCGATG	420	
Db	1458	TTGCCATGGCAATTAACAAAGCAAGGCTGAGACTCAGCAACCACTTGTGTTCATTCGATG	1399	
Qy	421	CAGGCTAGTAGTAAGTTTGGTTCCTGGTAGGAAAAAGGTCCTTATCTCACCCCTCTTAA	480	
Db	1398	CAGGCTAGTAGTAAGTTTGGTTCCTGGTAGGAAAAAGGTCCTTATCTCACCCCTCTTAA	1339	
Qy	481	ACTAAAGGTTCTTTTCAGGCT 500		
Db	1338	ACTAAAGGTTCTTTTCAGGCT 1319		
RESULT 5				
AF336376/c		2253 bp	mrna	linear
LOCUS				PRI 24-MAY-2001
DEFINITION				Homo sapiens platelet-derived growth factor D (PDGFD) mRNA,
				complete cds.
ACCESSION				AF336376
VERSION				AF336376.1
KEYWORDS				GI:14193795
SOURCE				Homo sapiens.
ORGANISM				Homo sapiens
REFERENCE				Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS				Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
				1 (bases 1 to 2253)
				Bergsten,E., Uetala,M., Li,X., Pietras,K., Ostman,A., Heidln,C.H.,
				Alitalo,K. and Eriksson,U.
TITLE				PDGF-D is a specific, protease-activated ligand for the PDGF
				beta-receptor
JOURNAL				Nat. Cell Biol. 3 (5), 512-516 (2001)
MEDLINE				21231379
PUBMED				11331881
REFERENCE				2 (bases 1 to 2253)
AUTHORS				Bergsten,E., Uetala,M., Li,X., Pietras,K., Ostman,A., Heidln,C.H.,
				Alitalo,K. and Eriksson,U.

Query Match	100.0%;	Score 500;	DB 9;	Length 2253;
Best Local Similarity	100.0%;	Pred. No. 5.8e-83;		
Matches 500;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	GTATATGTAAGAAGCCTCATCTTTTGATTTTTTAATATACAAGATGCTTCTTTTAAGAGA	60	
Db	1812	GTATATGTAAGAAGCCTCATCTTTTGATTTTTTAATATACAAGATGCTTCTTTTAAGAGA	1753	
Qy	61	GCAAGATTCAAAATTTGTTTGTCTTTTCAAAATTTAAAAATTAATTTATCTCCTTAAATTTT	120	
Db	1752	GCAAGATTCAAAATTTGTTTGTCTTTCAAAATTTAAAAATTAATTTATCTCCTTAAATTTT	1693	
Qy	121	CTAAGACATGTTTCATATATTTTGACATCCCTTATTTGGGCAAGAGATTTTAAAGAGTCT	180	
Db	1692	CTAAGACATGTTTCATATATTTTGACATCCCTTATTTGGGCAAGAGATTTTAAAGAGTCT	1633	
Qy	181	AACTCAAAACATATGTAAAGCTCTGGTGACCTGGTTATATATACCAAAAAAACAATTTGAT	240	
Db	1632	AACTCAAAACATATGTAAAGCTCTGGTGACCTGGTTATATATACCAAAAAAACAATTTGAT	1573	
Qy	241	CTATATACACATAGACATGAATATATTCTGTGTGTGTTGTGCATATATAAACCCTCAAAC	300	
Db	1572	CTATATACACATAGACATGAATATATTCTGTGTGTGTTGTGCATATATAAACCCTCAAAC	1513	
Qy	301	ACTATATTAAATGCAATCCCTATATTCTTAGGTATAGAGTTGATGATATACCTTTCTAC	360	
Db	1512	ACTATATTAAATGCAATCCCTATATTCTTAGGTATAGAGTTGATGATATACCTTTCTAC	1453	
Qy	361	TTGCCATGGCATTAACAAAGCAAGGCTGAGACTCAGCAACCACTTGTTGTCATTGCATTG	420	
Db	1452	TTGCCATGGCATTTAACAAGCAAGGCTGAGACTCAGCAACCACTTGTTGTCATTGCATTG	1393	
Qy	421	CAGGCTAGTAGTAAGTTTGGTTTGGTAGGAAAAGGGTCTCTATCTCACCCCTCCTTAA	480	
Db	1392	CAGGCTAGTAGTAAGTTTGGTTTGGTAGGAAAAGGGTCTCTATCTCACCCCTCCTTAA	1333	
Qy	481	ACTAAAGGTTCTTTTCAGGCT 500		
Db	1332	ACTAAAGGTTCTTTTCAGGCT 1313		


```
Qy 481 ACTAAGGTTCTTTTCAGGCT 500
Db 1252 ACTAAGGTTCTTTTCAGGCT 1233

RESULT 7
AX365305/c 3736 bp DNA linear PAT 15-FEB-2002
LOCUS
DEFINITION Sequence 1 from Patent W00189450.
ACCESSION AX365305
VERSION AX365305.1 GI:18697036
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Beals, J.M., Gonzalez-Dewhitt, P.A., Hammond, L.J., Lu, J., Na, S.,
Su, E.W., Witcher, D.R. and Wroblewski, V.J.
TITLE Cloning and characterization of the human gene for the protein
JOURNAL J. Biol. Chem. 274:11450-11455 (1999)
FEATURES
Location/Qualifiers
source 1..3736
/organism="Homo sapiens"
/db_xref="taxon:9606"
114..1226
/notes="unnamed protein product"
/codon_start=1
/protein_id="CAD23805.1"
/db_xref="GI:18697037"
/translation="MRLIFVYTLICANFCSCRDTSATPQASIKALRNANLRREDSE
HLNDLYRRDITQVKGNGYQSPFPNSYPRNLLTWRLHSQENTRIQLVFDNQFGLG
EANDLCRDYDEVEDISSTIIRGWCGRKEVPRIKSRNQIKITKSDDYFVAK
PGKIVYSLLEDQPAASSTNWSVTSSISGVSNPSVPTDPTLADALDKKIAEFD
TVEDLKYPNPSQEDLENMIDTPYRGRSHDRKSKVDLDRLNDKRYKRYCTPRN
YSNIEELKLANVFFPCLLVQRCGNGCGTVNWSCTCNSGKTVKRYHEVLQFE
PGHKRRGAKTALVDIQLDHERDCICSSRPPR"
sig_peptide 114..149
BASE COUNT 1164 a 736 c 718 g 1118 t

Query Match 100.0%; Score 500; DB 6; Length 3736;
Best Local Similarity 100.0%; Pred. No. 5,1e-83;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTATATGTAAGAAGCTCATCTTTGATTTTAAATATACAAAGATGCTTTCTTAAGAGA 60
Db 1750 GTATATGTAAGAAGCTCATCTTTGATTTTAAATATACAAAGATGCTTTCTTAAGAGA 1691

Qy 61 GCAAGATTCAAAATTTGTTTGTGTTTCAAAATTTAAATTTATCTCTCTAAATTTT 120
Db 1690 GCAAGATTCAAAATTTGTTTGTGTTTCAAAATTTAAATTTATCTCTCTAAATTTT 1631

Qy 121 CTAAGACATGTTTCATATATTTGACATCCCTTATTTTGGCAAGGATTTTAAGAGTCT 180
Db 1630 CTAAGACATGTTTCATATATTTGACATCCCTTATTTTGGCAAGGATTTTAAGAGTCT 1571

Qy 181 AACTCAACATATGTAAGCTCTGGTGTACCTGGTTATATATACAAATAAATTTGAT 240
Db 1570 AACTCAACATATGTAAGCTCTGGTGTACCTGGTTATATATACAAATAAATTTGAT 1511

Qy 241 CTATATACATACATAGATGAATATATTTCTGTGTGTTTGTGCATATATACCTCAAAAC 300
Db 1510 CTATATACATACATAGATGAATATATTTCTGTGTGTTTGTGCATATATACCTCAAAAC 1451

Qy 301 ACTATTATTAAATGCAATCCTATATTTCTTAAAGTATAGAGTTGATGATATACCTTTCTAC 360
Db 1450 ACTATTATTAAATGCAATCCTATATTTCTTAAAGTATAGAGTTGATGATATACCTTTCTAC 1391

Qy 361 TTGCCATGGATTAAACAAGCAAGGCTGAGACTCAGCAACCACTTGTGTTCATGCTG 420
Db 1390 TTGCCATGGATTAAACAAGCAAGGCTGAGACTCAGCAACCACTTGTGTTCATGCTG 1331
```

```
Qy 421 CAGGCTAGTAGTAAGTTTGTCTGTAGAAAAGGCTCTTTATCTACCTCTCTAA 480
Db 1330 CAGGCTAGTAGTAAGTTTGTCTGTAGAAAAGGCTCTTTATCTACCTCTCTAA 1271

RESULT 8
BC030645/c 4070 bp mRNA linear PRI 21-MAY-2002
LOCUS
DEFINITION IMAGE: 4824526, mRNA, complete cds.
ACCESSION BC030645
VERSION BC030645.1 GI:21040416
KEYWORDS MGC.
SOURCE MGC.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4070)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (20-MAY-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgaps@remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
Contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 33 Row: i Column: 12
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 15451919.

FEATURES
Location/Qualifiers
source 1..4070
/organism="Homo sapiens"
/db_xref="LocusID:80310"
/db_xref="taxon:9606"
/clone="MGC:26867 IMAGE:4824526"
/tissue_type="Testis"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/notes="Vector: pBluescript"
458..1552
/codon_start=1
/product="spinal cord-derived growth factor-B"
/protein_id="AAH30645.1"
/db_xref="GI:21040417"
/translation="MRLIFVYTLICANFCSCRDTSATPQASIKALRNANLRREDLY
RDEDTIQVKGNGYQSPFPNSYPRNLLTWRLHSQENTRIQLVFDNQFGLGAEANDI
CRYDFVEDEISSTIIRGWCGRKEVPRIKSRNQIKITKSDDYFVAKPGFKIY
YSLLEDQPAASSTNWSVTSSISGVSNPSVPTDPTLADALDKKIAEFDVLDL
KYPNPSQEDLENMIDTPYRGRSHDRKSKVDLDRLNDKRYKRYCTPRNYSNIE
ELKLANVFFPCLLVQRCGNGCGTVNWSCTCNSGKTVKRYHEVLQFEFGHLKR
RGRKTMALVDIQLDHERDCICSSRPPR"
BASE COUNT 1227 a 843 c 831 g 1169 t

Query Match 100.0%; Score 500; DB 9; Length 4070;
```


* 130771 130870: gap of 100 bp
 * 130871 160345: contig of 29475 bp in length.

FEATURES

source

Location/Qualifiers
 1. .160345
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="11"
 /map="11"

/clone="RP11-475J2"

/clone_lib="RPC1-11 Human Male BAC"

misc_feature

1. .946
 /note="assembly_fragment"

clone_end:17

vector_side:right

misc_feature

1047. .2803

/note="assembly_fragment"

misc_feature

2904. .4474

/note="assembly_fragment"

misc_feature

4575. .8112

/note="assembly_fragment"

misc_feature

8213. .11442

/note="assembly_fragment"

misc_feature

11543. .14027

/note="assembly_fragment"

misc_feature

14128. .17482

/note="assembly_fragment"

misc_feature

17583. .20007

/note="assembly_fragment"

misc_feature

clone_end:SP6

vector_side:right

misc_feature

20108. .26548

/note="assembly_fragment"

misc_feature

26649. .34423

/note="assembly_fragment"

misc_feature

34524. .41559

/note="assembly_fragment"

misc_feature

41660. .50886

/note="assembly_fragment"

misc_feature

50987. .64203

/note="assembly_fragment"

misc_feature

64304. .75627

/note="assembly_fragment"

misc_feature

75728. .91802

/note="assembly_fragment"

misc_feature

91903. .110040

/note="assembly_fragment"

misc_feature

110141. .130770

/note="assembly_fragment"

misc_feature

130871. .160345

/note="assembly_fragment"

BASE COUNT 48193 a 30956 c 30439 g 49055 t 1702 others

ORIGIN

Query Match 100.0%; Score 500; DB 2; Length 160345;
 Best Local Similarity 100.0%; Pred. No. 2e-83;
 Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTATATGTAAAGAGCCTCATCTTTTGATTTTAAATATACAAGATGCTTCTTTTAAGAGA 60

DB 82195 GTATATGTAAAGAGCCTCATCTTTTGATTTTAAATATACAAGATGCTTCTTTTAAGAGA 82136

QY 61 GCAAGATTCAAAATTTGTTTGTCTTCAAAATTTAAATTTTAAATTTTATCTCTCTAAATTTT 120

DB 82135 GCAAGATTCAAAATTTGTTTGTCTTCAAAATTTAAATTTTAAATTTTATCTCTCTAAATTTT 82076

QY 121 CTAAGACATGTTTCATATATTGACCATCCCTTTATTTGGCAAGAGATTTTAAGAGTCT 180

DB 82075 CTAAGACATGTTTCATATATTGACCATCCCTTTATTTGGCAAGAGATTTTAAGAGTCT 82016

QY 181 AACTCAACATATGTAAGTCTGCTGCTACCTGCTTATATATACCAAAAAACATTTGAT 240

DB 82015 AACTCAACATATGTAAGTCTGCTGCTACCTGCTTATATATACCAAAAAACATTTGAT 81956

QY 241 CTATATACATAGACATGAATATATTTCTGTGTGTGTGTGTGCATATATAACCTCAAAAC 300

DB 81955 CTATATACATAGACATGAATATATTTCTGTGTGTGTGTGTGCATATATAACCTCAAAAC 81896

QY 301 ACTATTATTAAATGCAATCCTATATTTCTTAGGTATAGAAGTTGATATATACCTTTCTAC 360

DB 81895 ACTATTATTAAATGCAATCCTATATTTCTTAGGTATAGAAGTTGATATATACCTTTCTAC 81836

QY 361 TTGCCATGGCATTACAAAAGAGGCTGAGACTCAGCAACCACTTGTGTTCATTGCATTG 420

DB 81835 TTGCCATGGCATTACAAAAGAGGCTGAGACTCAGCAACCACTTGTGTTCATTGCATTG 81776

QY 421 CAGGCTAGTAGTAAGTTTGTGCTGGTAGGAAAGGGTCTCTTATCTCACCCCTCTTAA 480

DB 81775 CAGGCTAGTAGTAAGTTTGTGCTGGTAGGAAAGGGTCTCTTATCTCACCCCTCTTAA 81716

QY 481 ACTAAAGGTTCTTTTCAGGCT 500

DB 81715 ACTAAAGGTTCTTTTCAGGCT 81696

RESULT 10

AP003043

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches 500;

Conservative

0; Mismatches

0; Indels

0; Gaps

0;

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY 181 AACTCAACATATGTAAGCTCTGGTACTGCTATATATATACCAAAAAACATTGTAT 240
|||||
Db 40043 AACTCAACATATGTAAGCTCTGGTACTGCTATATATATACCAAAAAACATTGTAT 40102
QY 241 CTATATACATAGACATGAATATATTTCTGTGTGTTGTCATATATACCTCAAC 300
|||||
Db 40103 CTATATACATAGACATGAATATATTTCTGTGTGTTGTCATATATACCTCAAC 40162
QY 301 ACTATTATTAATCAATCCTATATTTCTAGGTAGAGAGTTGATGATATACCTTTCTAC 360
|||||
Db 40163 ACTATTATTAATCAATCCTATATTTCTAGGTAGAGAGTTGATGATATACCTTTCTAC 40222
QY 361 TTGCCATGCATTAAACAGCAAGGCTGAGACTACGCAACACCTTGTGTTCAATG 420
|||||
Db 40223 TTGCCATGCATTAAACAGCAAGGCTGAGACTACGCAACACCTTGTGTTCAATG 40282
QY 421 CAGGCTAGTAGTAAGTTGGTTGCTGGTAGGAAAGGGTCTTATCTCACCTCTTAA 480
|||||
Db 40283 CAGGCTAGTAGTAAGTTGGTTGCTGGTAGGAAAGGGTCTTATCTCACCTCTTAA 40342
QY 481 ACTAAAGGTTCTTTTCAGGCT 500
|||||
Db 40343 ACTAAAGGTTCTTTTCAGGCT 40362

RESULT 11
AC024052
LOCUS
DEFINITION AC024052 182403 bp DNA linear HTG 17-AUG-2000
SEQUENCE, 7 unordered pieces.
AC024052
VERSION AC024052.3 GI:9838295
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 182403)
AUTHORS Waterston,R.H.
JOURNAL The sequence of Homo sapiens clone
2 (bases 1 to 182403)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (20-FEB-2000) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Aug 17, 2000 this sequence version replaced gi:7109658.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0617803
----- Summary Statistics -----
Sequencing vector: M13; 100%
Chemistry: Dye-terminator Big Dye; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Sequencing vector: plasmid; 0%
Assembly program: Phrap; version 0.990319
Consensus quality: 178574 bases at least Q40
Consensus quality: 179480 bases at least Q30
Consensus quality: 179940 bases at least Q20
Insert size: 198000; agarose-fp
Insert size: 183143; sum-of-ctg
Quality coverage: 5.39 in Q20 bases; agarose-fp
Quality coverage: 5.87 in Q20 bases; sum-of-ctg

* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1666: contig of 1666 bp in length
* 1667: gap of unknown length
* 1767: contig of 7367 bp in length
* 9134: gap of unknown length
* 9234: contig of 11430 bp in length
* 20663: gap of unknown length
* 20664: contig of 15164 bp in length
* 20764: gap of unknown length
* 35927: contig of 25583 bp in length
* 35928: gap of unknown length
* 36028: contig of 34457 bp in length
* 61611: gap of unknown length
* 61711: contig of 34457 bp in length
* 96168: gap of unknown length
* 96268: contig of 86136 bp in length.

FEATURES
source

1..182403
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/clone="RP11-617B3"
1..1666
/note="assembly_name:Contig2"
1767..9133
/note="assembly_name:Contig3"
clone_end:SP6
vector_side:left
9234..20663
/note="assembly_name:Contig4"
20764..35927
/note="assembly_name:Contig5"
36028..61610
/note="assembly_name:Contig6"
61711..96167
/note="assembly_name:Contig7"
96268..182403
/note="assembly_name:Contig8"
clone_end:T7
vector_side:left
55923 a 35190 c 34519 g 56169 t 602 others
ORIGIN

Query Match 100.0%; Score 500; DB 2; Length 182403;
Best Local Similarity 100.0%; Pred. No. 1.9e-83;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTATATGTAAGAAAGCCCTCATCTTTTGATTTTAAATATACAGATGCTTTCTTTAAGAGA 60
|||||
Db 136115 GTATATGTAAGAAAGCCCTCATCTTTTGATTTTAAATATACAGATGCTTTCTTTAAGAGA 136174
QY 61 GCAAGATTCAAAATGTTTGTGTTTCAAAATTTAAATTAATTTATCTCTAAATTTT 120
|||||
Db 136175 GCAAGATTCAAAATGTTTGTGTTTCAAAATTTAAATTAATTTATCTCTAAATTTT 136234
QY 121 CTAAGACATGTTTCATATATTTTGACCATCCCTTTATTTTGGCAAGAGATTTTAAGAGTCT 180
|||||
Db 136235 CTAAGACATGTTTCATATATTTTGACCATCCCTTTATTTTGGCAAGAGATTTTAAGAGTCT 136294
QY 181 AACTCAAAACATATGTAAGCTCTGTTGTTACCTGGTATATATACCAAAAAACATTTGAT 240
|||||
Db 136295 AACTCAAAACATATGTAAGCTCTGTTGTTACCTGGTATATATACCAAAAAACATTTGAT 136354
QY 241 CTATATACATAGACATGAATATATTTCTGTGTGTTGTCATATATACCTCAAC 300
|||||
Db 136355 CTATATACATAGACATGAATATATTTCTGTGTGTTGTCATATATACCTCAAC 136414
QY 301 ACTATTATTAATGCAATCCCTATATTTCTAGGTAGAGAGTTGATGATATACCTTTCTAC 360
|||||
Db 136415 ACTATTATTAATGCAATCCCTATATTTCTAGGTAGAGAGTTGATGATATACCTTTCTAC 136474

```

QY 361 TTGCCATGCATTACAAAGCAAGGCTGAGACTCAGCAACCACTGTGTTCATTGCATTG 420
Db 136475 TTGCCATGCATTACAAAGCAAGGCTGAGACTCAGCAACCACTGTGTTCATTGCATTG 136534
QY 421 CAGGCTAGTAGTAAGTGTGGTGGTAGGAAAAGGCTCTTATCTACCCCTCTTAA 480
Db 136535 CAGGCTAGTAGTAAGTGTGGTGGTAGGAAAAGGCTCTTATCTACCCCTCTTAA 136594
QY 481 ACTAAAGGTTCTTTCAGGCT 500
Db 136595 ACTAAAGGTTCTTTCAGGCT 136614

RESULT 12
AY027518/c 3710 bp mRNA linear PRI 17-JUL-2002
LOCUS Homo sapiens iris-expressed growth factor short form (IEGF) mRNA,
DEFINITION complete cds, alternatively spliced.
ACCESSION AY027518
VERSION AY027518.1 GI:13432062
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3710)
AUTHORS Wistow,G., Berstein,S.L., Ray,S., Wyatt,M.K., Behal,A.,
Touchman,J.W., Bouffard,G., Smith,D. and Peterson,K.
TITLE Expressed sequence tag analysis of adult human iris for the NEIBank
Project: steroid-response factors and similarities with retinal
pigment epithelium
JOURNAL Mol. Vis. 8, 185-195 (2002)
MEDLINE 22103462
PUBMED 12107412
REFERENCE 2 (bases 1 to 3710)
AUTHORS Wistow,G.
TITLE Direct Submission
JOURNAL Submitted (13-FEB-2001) MSF, NEI, 6/331, NIH, Bethesda, MD
20892-2740, USA
FEATURES
Source Location/Qualifiers
1..3710
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q22.3"
/tissue_type="iris"
1..3710
/gene="IEGF"
93..1187
/gene="IEGF"
/note="similar to PDGF/VEGF family; alternatively spliced"
/codon_start=1
/product="iris-expressed growth factor short form"
/protein_id="AAK20082.1"
/db_xref="GI:13432063"
/translation="MHRILFYVYLICANFCSCRDTSATPQSAIKALRNANLRDLDY
RRDETQVKGNGYVQSPNPRNLLTWRLHSONTRIQLVDFNOFGLEAENDI
CRYDEVEDISESTIIIRGWCGHKEVPRIKSRNIOIKITFKSDDYEVAKPGPKIY
YSLDFQPAASENWESVTSSISGVSNPSVDPLIADALDKKTAEDTVEDLL
KYFNPSQWEDLENMILDTPIRGRSYHDKRSKVDLRINDDAKRYSCTPRPNYSNIR
BELKIANVFFPRCLLVORCGNGCGGVNWRSCNCSGKTVKRYHEVLQEPGHIKR
RGRKTMALVDIOLDHRCDCICSSRRP"
BASE COUNT 1169 a 724 c 703 g 1114 t
ORIGIN

Query Match 99.4%; Score 496.8; DB 9; Length 3710;
Best Local Similarity 99.6%; Pred. No. 2e-82;
Matches 498; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTATATGAAGAAAGCCCTCATCTTTTGATTTTATATACAGATGCTTCTTTAAGAGA 60
Db 1711 GTATATGAAGAAAGCCCTCATCTTTTGATTTTATATACAGATGCTTCTTTAAGAGA 1652

```

```

QY 61 GCAAGATTCAAATAATGTTTGTGTTTCAAAAATTTAAAAATAAATTTATCTCTAAATTTT 120
Db 1651 GCAAGATTCAAATAATGTTTGTGTTTCAAAAATTTAAAAATAAATTTATCTCTAAATTTT 1592
QY 121 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTGGCAAGGATTTTAAAGAGTCT 180
Db 1591 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTGGCAAGGATTTTAAAGAGTCT 1532
QY 181 AACTCAACATATGTAAGCTCTGGTGTACCTGGTATATATACCAAAAAACATTTGAT 240
Db 1531 AACTCAACATATGTAAGCTCTGGTGTACCTGGTATATATACCAAAAAACATTTGAT 1472
QY 241 CTATATACATAGACATGAATATATTTCTGTGTGTGTTGTGCATATATACCTCAAC 300
Db 1471 CTATATACATAGACATGAATATATTTCTGTGTGTGTTGTGCATATATACCTCAAC 1412
QY 301 ACTATTTAAATGCAATCCCTATATTTCTAGGTATAGAAAGTTCATACCTTCTAC 360
Db 1411 ACTATTTAAATGCAATCCCTATATTTCTAGGTATAGAAAGTTCATACCTTCTAC 1352
QY 361 TTGCCATGCATTACAAAGCAAGGCTGAGACTCAGCAACCACTGTGTTCATTGCATTG 420
Db 1351 TTGCCATGCATTACAAAGCAAGGCTGAGACTCAGCAACCACTGTGTTCATTGCATTG 1292
QY 421 CAGGCTAGTAGTAAGTGTGGTGGTAGGAAAAGGCTCTTATCTACCCCTCTTAA 480
Db 1291 CAGGCTAGTAGTAAGTGTGGTGGTAGGAAAAGGCTCTTATCTACCCCTCTTAA 1232
QY 481 ACTAAAGGTTCTTTCAGGCT 500
Db 1231 ACTAAAGGTTCTTTCAGGCT 1212

RESULT 13
AY027517/c 3729 bp mRNA linear PRI 17-JUL-2002
LOCUS Homo sapiens iris-expressed growth factor long form (IEGF) mRNA,
DEFINITION complete cds, alternatively spliced.
ACCESSION AY027517
VERSION AY027517.1 GI:13432060
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3729)
AUTHORS Wistow,G., Berstein,S.L., Ray,S., Wyatt,M.K., Behal,A.,
Touchman,J.W., Bouffard,G., Smith,D. and Peterson,K.
TITLE Expressed sequence tag analysis of adult human iris for the NEIBank
Project: steroid-response factors and similarities with retinal
pigment epithelium
JOURNAL Mol. Vis. 8, 185-195 (2002)
MEDLINE 22103462
PUBMED 12107412
REFERENCE 2 (bases 1 to 3729)
AUTHORS Wistow,G.
TITLE Direct Submission
JOURNAL Submitted (13-FEB-2001) MSF, NEI, 6/331, NIH, Bethesda, MD
20892-2740, USA
FEATURES
Source Location/Qualifiers
1..3729
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q22.3"
/tissue_type="iris"
1..3729
/gene="IEGF"
93..1205
/gene="IEGF"
/note="similar to PDGF/VEGF family; alternatively spliced"
/codon_start=1
/product="iris-expressed growth factor long form"

```

/protein_id="AAK20081.1"
/db_xref="GI:13432061"
/translation="MRLIFVYTLICANFCSCRDTSATPOSASIKALRNANLRDEN
HLTDLYRDRDTIQVKGNGYQSPFNPSYPRNLLTWRLHSEQNTIQLVDFNOFGL
EAENDICRYDFVEVEDISESTIIRGWCCHKVPPRIKSRITQIKITFKSDDYFYAK
PGKIYVSLLEDFQPAASATNNESVTSSISGVSNPSVTDPLIADALDKKIAEFD
TVEDLLKYFNPESQWEDLENNYLDTPRYGRSHDRKSKVDLDLNDADKRYSCTPN
YSNIREELKLANVFFPRCLLVQRCGNGCGTVMNRSCNSGKTVMKRYKHYEVLQFE
PGHKRRGRAKTALVDIQLDHHERCDCICSSRPPR"
BASE COUNT 1177 a 730 c 706 g 1116 t
ORIGIN

Query Match 99.4%; Score 496.8; DB 9; Length 3729;
Best Local Similarity 99.6%; Pred. No. 2e-82;
Matches 498; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTATATCTAAGAAGCCCTCATCTTTTTCATTTTAAATATACAAGATGCTTTCTTTAAGAGA 60
|||||
Db 1729 GTATATCTAAGAAGCCCTCATCTTTTTCATTTTAAATATACAAGATGCTTTCTTTAAGAGA 1670
|||||
Qy 61 GCAAGATTCAAAATGTTTGTGTTTCAAAATTTTAAATATACAAGATGCTTTCTTTAAGAGA 120
|||||
Db 1669 GCAAGATTCAAAATGTTTGTGTTTCAAAATTTTAAATATACAAGATGCTTTCTTTAAGAGA 1610
|||||
Qy 121 CTAAGACATGTTTCATATATTGACCATCCCTTATTTGGCAAGAGATTTTAAGAGTCT 180
|||||
Db 1609 CTAAGACATGTTTCATATATTGACCATCCCTTATTTGGCAAGAGATTTTAAGAGTCT 1550
|||||
Qy 181 AACTCAACATATGTAAGCTCTGCTGACCTGGTATATATACCAAAAACATTTGAT 240
|||||
Db 1549 AACTCAACATATGTAAGCTCTGCTGACCTGGTATATATACCAAAAACATTTGAT 1490
|||||
Qy 241 CTATATACATAGACATGAATATATTTCTGCTGCTGTTTGTGCATATATAACCTCAAC 300
|||||
Db 1489 CTATATACATAGACATGAATATATTTCTGCTGCTGTTTGTGCATATATAACCTCAAC 1430
|||||
Qy 301 ACTATTATTAATGCAATCTATATCTTAGGTATAGAAGTTGATGATATACCTTTCTAC 360
|||||
Db 1429 ACTATTATTAATGCAATCTATATCTTAGGTATAGAAGTTGATGATATACCTTTCTAC 1370
|||||
Qy 361 TTGCATGGCATTAACAAGCAAGGCTGAGACTCAGCAACCACTTGTTGTTCAATGCAATG 420
|||||
Db 1369 TTGCATGGCTTTAACAAGCAAGGCTGAGACTCAGCAACCACTTGTTGTTCAATGCAATG 1310
|||||
Qy 421 CAGCTAGTAGTAGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 480
|||||
Db 1309 CAGCTAGTAGTAGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1250
|||||
Qy 481 ACTAAGGTTCTTTTCAGGCT 500
|||||
Db 1249 ACTAAGGTTCTTTTCAGGCT 1230
|||||

RESULT 14
AX044487/c
LOCUS AX044487 1882 bp DNA linear PAT 24-NOV-2000
DEFINITION Sequence 1 from Patent WO0066736.
ACCESSION AX044487
VERSION AX044487.1 GI:11343345
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1882)
AUTHORS Gilbert, T., Hart, C.E., Sheppard, P.O. and Gilbertson, D.G.
TITLE Growth factor homolog zveg4
JOURNAL Patent: WO 0066736-A 1 09-NOV-2000;
ZymoGenetics, Inc. (US)
FEATURES
source
1. .1882
/organism="Homo sapiens"
/db_xref="taxon:9606"

CDS 226...1338
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC17169.1"
/db_xref="GI:11343346"
/translation="MRLIFVYTLICANFCSCRDTSATPOSASIKALRNANLRDEN
HLTDLYRDRDTIQVKGNGYQSPFNPSYPRNLLTWRLHSEQNTIQLVDFNOFGL
EAENDICRYDFVEVEDISESTIIRGWCCHKVPPRIKSRITQIKITFKSDDYFYAK
PGKIYVSLLEDFQPAASATNNESVTSSISGVSNPSVTDPLIADALDKKIAEFD
TVEDLLKYFNPESQWEDLENNYLDTPRYGRSHDRKSKVDLDLNDADKRYSCTPN
YSNIREELKLANVFFPRCLLVQRCGNGCGTVMNRSCNSGKTVMKRYKHYEVLQFE
PGHKRRGRAKTALVDIQLDHHERCDCICSSRPPR"
BASE COUNT 566 a 407 c 430 g 479 t
ORIGIN

Query Match 96.6%; Score 482.8; DB 6; Length 1882;
Best Local Similarity 98.8%; Pred. No. 9e-80; 2; Indels 4; Gaps 1;
Matches 498; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

Qy 1 GTATATCTAAGAAGCCCTCATCTTTTTCATTTTAAATATACAAGATGCTTTCTTTAAGAGA 60
|||||
Db 1866 GTATATCTAAGAAGCCCTCATCTTTTTCATTTTAAATATACAAGATGCTTTCTTTAAGAGA 1807
|||||
Qy 61 GCAAGATTCAAAATGTTTGTGTTTCAAAATTTTAAATATACAAGATGCTTTCTTTAAGAGA 120
|||||
Db 1806 GCAAGATTCAAAATGTTTGTGTTTCAAAATTTTAAATATACAAGATGCTTTCTTTAAGAGA 1747
|||||
Qy 121 CTAAGACATGTTTCATATATTGACCATCCCTTATTTGGCAAGAGATTTTAAGAGTCT 180
|||||
Db 1746 CTAAGACATGTTTCATATATTGACCATCCCTTATTTGGCAAGAGATTTTAAGAGTCT 1687
|||||
Qy 181 AACTCAACATATGTAAGCTCTGCTGACCTGGTATATATATAC - - - AAAAAAACATT 236
|||||
Db 1686 AACTCAACATATGTAAGCTCTGCTGACCTGGTATATATATACCAAAAAAACATT 1627
|||||
Qy 237 TGATCTATATACATAGACATGAATATATTTCTGCTGTTTGTGCATATATAACCTC 296
|||||
Db 1626 TGATCTATATACATAGACATGAATATATTTCTGCTGTTTGTGCATATATAACCTC 1567
|||||
Qy 297 AAACACTATTATTAATGCAATCTATATTTCTTAGGTATAGAAGTTTCATATACCTTT 356
|||||
Db 1566 AAACACTATTATTAATGCAATCTATATTTCTTAGGTATAGAAGTTTCATATACCTTT 1507
|||||
Qy 357 CTACTTGCCATGGCATTAAACAAAGAGGCTGAGACTCAGCAACCACTTGTTGTTCAATGC 416
|||||
Db 1506 CTACTTGCCATGGCATTAAACAAAGAGGCTGAGACTCAGCAACCACTTGTTGTTCAATGC 1447
|||||
Qy 417 ATTGCAGGCTAGTAGTAAGTTTGGTCTGCTGGTAGGAAAAGGCTCTTATCTCACCTCC 476
|||||
Db 1446 ATTGCAGGCTAGTAGTAAGTTTGGTCTGCTGGTAGGAAAAGGCTCTTATCTCACCTCC 1387
|||||
Qy 477 TTAACACTAAAGGTTCTTTTCAGGCT 500
|||||
Db 1386 TTAACACTAAAGGTTCTTTTCAGGCT 1363
|||||

RESULT 15
AX207483/c
LOCUS AX207483 1882 bp DNA linear PAT 31-AUG-2001
DEFINITION Sequence 1 from Patent WO0157083.
ACCESSION AX207483
VERSION AX207483.1 GI:15422211
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1882)
AUTHORS Gilbertson, D.G. and Hart, C.E.
TITLE Methods for promoting growth of bone, ligament, and cartilage using zveg4
JOURNAL Patent: WO 0157083-A 1 09-AUG-2001;
ZymoGenetics, Inc. (US)

FEATURES
source Location/Qualifiers
1..1882
/organism="Homo sapiens"
/db_xref="taxon:9606"
226..1338
CDS
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC60334.1"
/db_xref="GI:15422212"
/translation="MRLIFVVTLICANFCSCRDTSATPQASIKALRNANLRDSEN
HLDLYRRDETIOVKGNGYVQSPRPNSYPNLLLTWRLHSENRPIQLVFNQGLE
EANDICRDFVEDEIDSETSLIIRGWGHEVPPIKSRNTQIKRTPKSDDIYFAK
PGFKIYSLLEDFQAAASETNWSSISGVSNPSVPTDPTLIADALDKKIAEFD
TVEDLLKYFNPSQWEDLENMYLDTPRYGRSKVDRKSKVDLRDLNDADAKRYSCTPRN
YSVNTIREELKLANVFFPRCLLVQRCGNGCGGTWNRSCTCNSGKTVKKYHEVLQFE
PGHKRRGRKTMALVDIOLDHHERCDICSSRPPR"

BASE COUNT 566 a 407 c 430 g 479 t
ORIGIN

Query Match 96.6%; Score 482.8; DB 6; Length 1882;
Best Local Similarity 98.8%; Pred. No. 9e-80;
Matches 498; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

QY 1 GTATATGTAAAGAGCCTCATCTTTTGATTTTAAATATACAGATGCTTCTTTAAAGACA 60
Db |||||||
QY 61 GCAAGATTCAAAATCTTTTGGTTTCAAAATTTAAATAAATTTATCTCTAAATTTT 120
Db |||||||
QY 1806 GCAAGATTCAAAATCTTTTGGTTTCAAAATTTAAATAAATTTATCTCTAAATTTT 1747
QY 121 CTAAGACATGTTTCATATATTGACCATCCCTTATTTGGCAAGGATTTTAAAGAGTCT 180
Db |||||||
QY 1746 CTAAGACATGTTTCATATATTGACCATCCCTTATTTGGCAAGGATTTTAAAGAGTCT 1687
QY 181 AACTCAACATATGTAAGCTCTGGTGACCTGCTGTTATATATACC-----AAAAAACATT 236
Db |||||||
QY 1686 AACTCAACATATGTAAGCTCTGGTGACCTGCTGTTATATATACC-----AAAAAACATT 1627
QY 237 TGATCTATATACATAGACATGAATATATTCTGTGCTGTTTGGCATATATACCTC 296
Db |||||||
QY 1626 TGATCTATATACATAGACATGAATATATTCTGTGCTGTTTGGCATATATACCTC 1567
QY 297 AAACACTATTATAATGCAATCCTATATTCTAGGTATAGAAGTTGATGATATACCTTT 356
Db |||||||
QY 1566 AAACACTATTATAATGCAATCCTATATTCTAGGTATAGAAGTTGATGATATACCTTT 1507
QY 357 CTACTTGCATGCGCATTAACAAGCAAGCGCTGAGACTCAGCAACCACTTGTGTTCAATGC 416
Db |||||||
QY 1506 CTACTTGCATGCGCATTAACAAGCAAGCGCTGAGACTCAGCAACCACTTGTGTTCAATGC 1447
QY 417 ATTGCAGGCTAGTAGTAAGTTGGTTGCTGGTAGGAAGGGTCTCTTATCTCACCTCC 476
Db |||||||
QY 1446 ATTGCAGGCTAGTAGTAAGTTGGTTGCTGGTAGGAAGGGTCTCTTATCTCACCTCC 1387
QY 477 TTAACCTAAAGGTTCTTTCAGGCT 500
Db |||||||
QY 1386 TTAACCTAAAGGTTCTTTCAGGCT 1363

Search completed: July 4, 2003, 16:40:37
Job time : 1509.66 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 15:04:06 ; Search time 166.445 seconds

(without alignments)

6765.003 Million cell updates/sec

US-10-083-853B-2_COPY_1_500

Perfect score:

Sequence: 1 gtatatgtaagaaagcctca.....actaaagggttcttccaggct 500

Scoring table: IDENTITY NIC

IDENTITY_NOC
Gapop 10.0 ; Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length:	0
Maximum DB seq length:	2000000000

Post-processing: Minimum Match 0%

100% processing. Minimum Match 0%
Maximum Match 100%

-----Listing first-45-summaries

```

Database: Genesec. 101002.*
1: /SID$2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SID$2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SID$2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SID$2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SID$2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SID$2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SID$2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SID$2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SID$2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SID$2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SID$2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SID$2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SID$2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SID$2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SID$2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SID$2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SID$2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SID$2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SID$2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SID$2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SID$2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SID$2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SID$2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SID$2/gcgdata/geneseq/geneseqn-emb1/NA2001AT.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARY

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
C 1	500	100.0		1587	22	AAS04493	Human FCTR2 DNA pr
C 2	500	100.0		1728	22	AAS04498	Human PDGFD DNA
C 3	500	100.0		1828	21	AAS04492	Human FCTR1 DNA pr
C 4	500	100.0		1934	21	AAD00737	Human Platelet Der
C 5	500	100.0		2253	21	AAD00738	Human Platelet Der
C 6	500	100.0		3718	24	AAD25488	Human LP85 DNA #2.
C 7	500	100.0		3736	24	AAD25489	Human LP85 DNA #1.
C 8	499.6	99.9		4001	22	AAH46957	Human secreted pro
C 9	498.4	99.7		3798	22	AAH46939	Human secreted pro

10	497.6	99.5	2726	22	AAH646959	Human secreted protein
C 11	484.8	97.0	3853	22	AAF241396	Human VEGF-G CDNA.
C 12	482.8	96.6	1882	21	AAC931555	Human growth factor
C 13	485.1	96.6	1882	21	AAAS1541	SEQ. ID. 36 from W
C 14	482.8	96.6	1882	22	AAH47772	Human zveg44 polypp
C 15	482.8	96.6	1882	22	AAH47773	Human zveg44 polypp
C 16	460	92.0	485	24	ABQ73239	Human zveg44 polypp
C 17	131.8	26.4	360	21	ABL63632	Breast cancer rela
C 18	131.8	26.4	620	22	AAD007739	Human platelet der
C 19	109.2	21.8	3121	22	ABA08941	Human novel protei
C 20	74.2	14.8	1472	21	AAF241398	Human VEGF-G CDNA.
C 21	74.2	14.8	1472	22	AAC841596	Mouse growth facto
C 22	74.2	14.8	1472	22	AAH47773	Mouse zveg44 polypp
C 23	62	12.4	690	21	ABQ73240	Mouse zveg44 polypp
C 24	61.4	12.3	256	22	AAD007736	Human platelet der
C 25	61.4	12.3	256	22	AAAS46959	Human breast cance
C 26	61.4	12.3	256	24	AAF17529	Human breast cance
C 27	56.4	11.3	7536	24	ABK94995	Human breast tumor
C 28	51.4	10.3	6244	24	ABL333465	Human immune syste
C 29	51.4	10.3	7346	24	ABL332484	Human immune syste
C 30	51.4	10.3	9810	24	ABL32344	Human immune syste
C 31	51	10.2	6167	24	ABK32436	Human immune syste
C 32	49.6	9.9	14551	24	ABK28369	DNA transcription
C 33	49.4	9.9	19124	18	ABL34584	Human metastasis a
C 34	49.4	9.9	19124	18	AAAT72882	Plasmodium var-7 g
C 35	49.2	9.8	5883	21	AAZ58287	Plasmodium var-7 p
C 36	49.2	9.8	5883	24	ABL32352	Human immune syste
C 37	49.2	9.8	6641	24	ABL34472	Human metastasis a
C 38	49.2	9.8	6641	24	ABL54335	Chemically treated
C 39	49	9.8	229	24	ABL32314	Human immune syste
C 40	49	9.8	229	24	ABL63535	Breast cancer rela
C 41	49	9.8	6012	24	ABL63968	Breast cancer rela
C 42	49	9.8	6012	24	ABL70328	Chemically treated
C 43	49	9.8	6012	24	AAAG1275	Human gene regulat
C 44	49	9.8	6012	24	ABK31371	Signal transductio
C 45	49	9.8	9810	24	ABL32427	Human immune syste
C 46	49	9.8	18157	24	ABO67083	Human angiogenesis

ALIGNMENTS

RESULT 1

RESOLUT
AAS04493/C

ID AAS04493 standard: DNA: 1587 BP.

AX
AC

DT 07-SEP-2001 (first entry)

XX
DE
Human ECTP2 DNA present in clone 30664188 0 331

Bone morphogenetic protein-1; BMP-1; vascular endothelial growth factor; VEGF-E; platelet derived growth factor; PDGF; hyperplasia; cancer; neoplasia; anemia; leukopenia; baldness; cardiovascular disorder; fibrotic disorder; diabetic ulcer; obesity; hyperproliferation; human; dysproliferation; neurodegenerative disorder; osteoarthritis; epilepsy; inflammatory disorder; Graft versus host disease; coagulation; ds; haemophilia; neural disorder; Parkinson's disease; Alzheimer's disease; multiple sclerosis; Huntington's disease; amyotrophic lateral sclerosis; peripheral neuropathy; acute brain injury.

Homo sapiens.

[illegible]

key	LOCATION
CDS	540 938

```

CDS
C40.1.530
/*tag= a

```

```

--/product="Human ECTR2"

```

W0200125437-A2

9ZAY#K6E2T0020M

12-APR-2001

10025479-07

PR	08-AUG-2000; 2000US-0223879.
PR	12-SEP-2000; 2000US-0662783.
PR	20-SEP-2000; 2000US-0234082.
XX	(CURA-) CURAGEN CORP.
PA	
PI	Shinkets RA, Lichenstein H, Herrmann JL, Boldog FL, Minskoff S;
PI	Jeffers M;
XX	
DR	WPI: 2001-316172/33.
DR	P-PSDB; AAU00704.
XX	
PT	Novel growth factor polypeptides termed as FCTR polypeptides, useful
PT	for treating cancer, cardiovascular and fibrotic diseases, diabetic
PT	ulcers, wound healing and neuronal disorders -
XX	
PS	Disclosure; Fig 13; 17lpp; English.
XX	
CC	The sequence represents DNA encoding a protein related to bone
CC	morphogenetic protein-1 (BMP-1), vascular endothelial growth factor
CC	(VEGF-E) and platelet derived growth factor (PDGF). Polypeptides and
CC	polynucleotides related to BMP-1, VEGF-E and PDGF are referred to as
CC	FCTR peptides and nucleic acids. FCTR proteins are useful for treating
CC	or preventing a disorder associated with aberrant expression, aberrant
CC	processing, or aberrant physiological interactions of the proteins in a
CC	mammal, where the disorder is characterised by insufficient or
CC	ineffective growth of a cell or a tissue, e.g. hyperplasia or neoplasia.
CC	The peptides and their associated nucleic acids are useful for both
CC	promoting and inhibiting growth of cells and tissues and in treatment of
CC	cancer, anaemia, leukopenia, baldness, for treating cardiovascular and
CC	fibrotic disorders, diabetic ulcers, obesity, infectious diseases,
CC	hyperproliferative and dysproliferative disorders, neurodegenerative
CC	diseases, osteoarthritis, inflammatory disorders, Graft versus host
CC	disease, coagulation disorders such as haemophilia, and neural disorders
CC	including Parkinson's disease, Alzheimer's disease, multiple sclerosis,
CC	Huntington's disease, amyotrophic lateral sclerosis, peripheral
CC	neuropathy, acute brain injury and epilepsy.
XX	
SQ	Sequence 1728 BP; 530 A; 364 C; 379 G; 455 T; 0 other;
	Query Match 100.0%; Score 500; DB 22; Length 1728;
	Best Local Similarity 100.0%; Pred. No. 3e-96;
	Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 GTATGTCTAAGAAGCCTCATCTTTTCATTCTTAAATACAAGATGCTTCTTTTAAGAGA 60
Db	
1718	GTAFTGTGAAGAAGCCTCACTTTTGATTTTTTAATACAAGATGCTTCTTTTAAGAGA 1659
QY	61 GCAAGATTCAAAATGTTTGTGTTTCAAATTTAAAAATAAAATTTATCTCTCTAAATTTT 120
Db	
1658	GCAAGATTCAAAATGTTTGTGTTTCAAATTTAAAAATAAAATTTATCTCTCTAAATTTT 1599
QY	121 CTAAACACATGTTTCATATATTTTGACATCCCTTATTTGGCAAGAGATTTTAAGAGTCT 180
Db	
1598	CTAAACACATGTTTCATATATTTTGACATCCCTTATTTGGCAAGAGATTTTAAGAGTCT 1539
QY	181 AACCTCAAAACATATGTAAAGCTCTGGTGACTCGTTATATATACCACAAAAAACATTTGAT 240
Db	
1538	AACCTCAAAACATATGTAAAGCTCTGGTGACTCGTTATATATACCACAAAAAACATTTGAT 1479
QY	241 CTATATACATAGACATGAATATTTCTGTGTGTGTGTGTGCATATATAACCTCAAAC 300
Db	
1478	CTATATACATAGACATGAATATTTCTGTGTGTGTGTGTGCATATATAACCTCAAAC 1419
QY	301 ACTATTATTAATGCCATCCTATATTTCTTAGGTATAGAAGTTGATATACCTTTCTAC 360
Db	
1418	ACTATTATTAATGCCATCCTATATTTCTTAGGTATAGAAGTTGATATACCTTTCTAC 1359
QY	361 TTGCCATGGCATTAAACAAAGCAAGGCTTGACACTCAGCACCACTTGTTGTCATTGATTC 420
Db	
1358	TTGCCATGGCATTAAACAAAGCAAGGCTTGACACTCAGCACCACTTGTTGTCATTGATTC 1299
QY	421 CAGGCTAGTAGTAAGTTTGGTTGCTGGTAGGAAAAAGGGTCTCTATATCTACCCCTCTTAA 480

CC FCTR peptides and nucleic acids. FCTR proteins are useful for treating
CC or preventing a disorder associated with aberrant expression, aberrant
CC processing, or aberrant physiological interactions of the proteins in a
CC mammal, where the disorder is characterised by insufficient or
CC ineffective growth of a cell or a tissue, e.g. hyperplasia or neoplasia.
CC The peptides and their associated nucleic acids are useful for both
CC promoting and inhibiting growth of cells and tissues and in treatment of
CC cancer, anaemia, leukopenia, baldness, for treating cardiovascular and
CC fibrotic disorders, diabetic ulcers, obesity, infectious diseases,
CC hyperproliferative and dysproliferative disorders, neurodegenerative
CC disorders, osteoarthritis, inflammatory disorders, Graft versus host
CC disease, coagulation disorders such as haemophilia, and neural disorders
CC including Parkinson's disease, Alzheimer's disease, multiple sclerosis,
CC Huntington's disease, amyotrophic lateral sclerosis, peripheral
CC neuropathy, acute brain injury and epilepsy.
XX
XX Sequence 1828 BP; 572 A; 380 C; 385 G; 491 T; 0 other;

Query Match 100.0%; Score 500; DB 22; Length 1828;
Best Local Similarity 100.0%; Pred. No. 3.le-96;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTATATGTAAGAAAGCCCTCATCTTTTGTATTTTATATACAGATGCTTCTTTAAGAGA 60
Db 1818 GTATATGTAAGAAAGCCCTCATCTTTTGTATTTTATATACAGATGCTTCTTTAAGAGA 1759

QY 61 GCAAGATTCAAAATTTGTTTGTGTTTCAAAATTTTAAATTAATTTATCTCTAAATTTT 120
Db 1758 GCAAGATTCAAAATTTGTTTGTGTTTCAAAATTTTAAATTAATTTATCTCTAAATTTT 1699

QY 121 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAGAGATTTTAAAGTCT 180
Db 1698 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAGAGATTTTAAAGTCT 1639

QY 181 AACTCAACATATGTAAGCTCTGTTGTTACCTGTTTATATATACCAAAAAACATTTGAT 240
Db 1638 AACTCAACATATGTAAGCTCTGTTGTTACCTGTTTATATATACCAAAAAACATTTGAT 1579

QY 241 CTATATACATAGACATGAATATATTTCTGTGTTGTTTGTGCATATATACCTCAAC 300
Db 1578 CTATATACATAGACATGAATATATTTCTGTGTTGTTTGTGCATATATACCTCAAC 1519

QY 301 ACTATATTAATGCAATCCATATATCTTAGTATAGAGTATGATATACCTTTCTAC 360
Db 1518 ACTATATTAATGCAATCCATATATCTTAGTATAGAGTATGATATACCTTTCTAC 1459

QY 361 TTGCGATGCAATTAACAAGCAAGCTGAGACTCAGCAACCACTTGTGTTCAATTCATTG 420
Db 1458 TTGCGATGCAATTAACAAGCAAGCTGAGACTCAGCAACCACTTGTGTTCAATTCATTG 1399

QY 421 CAGGCTAGTAGTAAGTTTGGTTGTTGTTAGGAAAGGCTCTTATCTCACCCCTCTTAA 480
Db 1398 CAGGCTAGTAGTAAGTTTGGTTGTTGTTAGGAAAGGCTCTTATCTCACCCCTCTTAA 1339

QY 481 ACTAAGGTTCTTTTCAGGCT 500
Db 1338 ACTAAGGTTCTTTTCAGGCT 1319

Result: 4
AAND00737/c

ID AAND00737 standard; cDNA; 1934 BP.
XX
AC AAD00737;
XX
DT 08-SEP-2000 (first entry)
XX
DE Human Platelet Derived Growth Factor (PDGF)-D encoding partial cDNA #2.
XX
KW Platelet Derived Growth Factor-D; PDGF-D; human; cytostatic; vulnary;
KW VEGF-G; Vascular Endothelial Growth Factor; antiatherosclerotic; tumour;
KW proliferative; activator; proliferation; differentiation; motility;
KW growth; PDGF-D receptor; antagonist; tissue remodelling; treat;

KW atherosclerosis; wound; metastasis; ss.
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..969
FT /partial
FT /tag- a
FT /product= "Human PDGF-D partial polypeptide #2"
FT /note= "5' truncated platelet derived growth factor"
XX
XX WO200027879-A1.
XX 18-MAY-2000.
XX
PF 10-NOV-1999; 99WO-US26462.
PR 10-NOV-1998; 98US-0107852.
PR 28-DEC-1998; 98US-0113997.
PR 26-AUG-1999; 99US-0150604.
PR 04-OCT-1999; 99US-0157108.
PR 05-OCT-1999; 99US-0157756.
XX
XX (LUDW-) LUDWIG INST CANCER RES.
PA (UYHE-) UNIV HELSINKI LICENSING LTD OY.
XX
PI Eriksson U, Aase K, Ponten A, Lee X, Uutela M, Alitalo K;
PI Oestman A, Heldin C;
XX
DR WPI: 2000-376495/32.
DR P-PSDB; AAY71129.
XX
PT Novel polynucleotides encoding a novel growth factor of cells
PT expressing a platelet-derived growth factor, useful for diagnostic and
PT therapeutic applications, e.g. concerning cancer -
XX
XX Claim 1: Fig 5; 11lpp; English.

CC The present sequence is the 5' truncated partial cDNA #2, encoding human
CC platelet derived growth factor (PDGF)-D, formally known as Vascular
CC Endothelial Growth Factor (VEGF)-G. It is derived from human foetal lung
CC lambda210 cDNA library. It belongs to the VEGF/PDGF family. It functions
CC as an activator of proliferation, differentiation, growth and motility of
CC cells, that express PDGF-D receptor. This sequence is useful for
CC inhibiting the growth of tumours, that express PDGF-D. Expression of
CC PDGF-D and its proteolytic cleavage for generating an activated truncated
CC form is useful for regulating receptor binding specificity of PDGF-D.
CC PDGF-D antagonist is useful for inhibiting tissue remodelling during the
CC invasion of tumour cells into normal cells. PDGF-D may be used to treat
CC wounds, atherosclerosis, metastasis and migration of smooth muscle cells.
XX
XX Sequence 1934 BP; 632 A; 366 C; 394 G; 542 T; 0 other;

Query Match 100.0%; Score 500; DB 21; Length 1934;
Best Local Similarity 100.0%; Pred. No. 3.le-96;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTATATGTAAGAAAGCCCTCATCTTTTGTATTTTATATACAGATGCTTCTTTAAGAGA 60
Db 1493 GTATATGTAAGAAAGCCCTCATCTTTTGTATTTTATATACAGATGCTTCTTTAAGAGA 1434

QY 61 GCAAGATTCAAAATTTGTTTGTGTTTCAAAATTTTAAATTAATTTATCTCTAAATTTT 120
Db 1433 GCAAGATTCAAAATTTGTTTGTGTTTCAAAATTTTAAATTAATTTATCTCTAAATTTT 1374

QY 121 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAGAGATTTTAAAGTCT 180
Db 1373 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAGAGATTTTAAAGTCT 1314

QY 181 AACTCAACATATGTAAGCTCTGTTGTTACCTGTTTATATATACCAAAAAACATTTGAT 240
Db 1313 AACTCAACATATGTAAGCTCTGTTGTTACCTGTTTATATATACCAAAAAACATTTGAT 1254

|||||
Db 1699 ACTATTATTAATGCAATCTATATTCTTAGGTATAGAAGTTGATGATATACCTTCTAC 1640
QY 361 TTGCCATGGCATTTAAACAAGCAAGGCTGAGACTCAGCAACCACTTGTTGTTCAATGCAATG 420
Db 1639 TTGCCATGGCATTTAAACAAGCAAGGCTGAGACTCAGCAACCACTTGTTGTTCAATGCAATG 1580
QY 421 CAGGCTAGTAGTAAGTTTGGTTGCTGCTAGGAAAAGGCTCTATCTCACCCCTCCTTAA 480
Db 1579 CAGGCTAGTAGTAAGTTTGGTTGCTGCTAGGAAAAGGCTCTATCTCACCCCTCCTTAA 1520
QY 481 ACTAAAGGTTCTTTCAGGCT 500
Db 1519 ACTAAAGGTTCTTTCAGGCT 1500

RESULT 9
AAH46939/c
ID AAH46939 standard; cDNA; 3798 BP.
XX
AC AAH46939;
XX
DT 25-SEP-2001 (first entry)
XX
DE Human secreted protein encoding cDNA (clone Id HGCNC48).
XX
KW Secreted protein; immunosuppressive; antiarthritic; antirheumatic;
KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
KW nootropic; neuroprotective; antibacterial; virucide; fungicide; human;
KW ophthalmological; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN WO200155430-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01431.
XX
PP 31-JAN-2000; 2000US-0179065.
XX
PR 04-FEB-2000; 2000US-0180628.
XX
PS 12-SEP-2000; 2000US-0231968.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Komatsoulis GA, Baker KP, Birse CE, Soppet DR, Olsen HS;
PI Moore PA, Wei P, Ebner R, Duan DR, Shi Y, Choi GH, Fiscella M;
PI Ni J, Ruben SM, Barash SC;
XX
DR WPI; 2001-476220/51.
XX
PT P-PSDB; AAB85529.
XX
PT 17 isolated nucleic acid molecules encoding human secreted proteins,
XX used to preventing, treating or ameliorating a medical condition -
XX
PS Claim 1; Page 417-418; 482pp; English.
XX
XX The invention provides novel human secreted proteins and polynucleotides
XX encoding them. The secreted proteins can be expressed by standard
XX recombinant methodology. The secreted proteins and polynucleotides are
XX used to prevent, treat or ameliorate a medical condition in e.g. humans,
XX mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They can
XX also be used in diagnosing a pathological condition. The antibodies to
XX the proteins can also be used in alleviating symptoms associated with the
XX disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme
XX linked immunosorbent assays (ELISA). Disorders which are diagnosed or
XX treated include autoimmune diseases e.g. rheumatoid arthritis,
XX hyperproliferative disorders e.g. neoplasms of the breast or liver,
XX cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
XX e.g. cerebral ischemia, angiogenesis, nervous system disorders e.g.
XX Alzheimer's disease, infections caused by bacteria, viruses and fungi and
XX ocular disorders e.g. corneal infection. The polypeptides can also be
XX used to aid wound healing and epithelial cell proliferation, to prevent

CC skin aging due to sunburn, to maintain organs before transplantation, for
CC supporting cell culture of primary tissues, to regenerate tissues and in
CC chemotaxis. The polypeptides can also be used as a food additive or
CC preservative to increase or decrease storage capabilities. The present
CC sequence represents a human secreted protein encoding cDNA.
XX
SQ Sequence 3798 BP; 1189 A; 749 C; 737 G; 1123 T; 0 other;

Query Match 99.7%; Score 498.4; DB 22; Length 3798;
Best Local Similarity 99.8%; Pred. No. 7.1e-96;
Matches 499; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTATATGTAAGAAAGCCTCATCTTTTGATTTTAAATATACAAGATGCTTTCTTTAAGAGA 60
Db 1802 GTATATGTAAGAAAGCCTCATCTTTTGATTTTAAATATACAAGATGCTTTCTTTAAGAGA 1743
QY 61 GCAAGATTCAAAATTTGTTGTTTCAAAATTTAAATAAATTTATCTCTAAATTTT 120
Db 1742 GCAAGATTCAAAATTTGTTGTTTCAAAATTTAAATAAATTTATCTCTAAATTTT 1683
QY 121 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAGGATTTTAAAGAGTCT 180
Db 1682 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAGGATTTTAAAGAGTCT 1623
QY 181 AACTCAACATATGTAAGCTCTGGTGACCTGGTTATATATATACCAAAAAACATTTGAT 240
Db 1622 AACTCAACATATGTAAGCTCTGGTGACCTGGTTATATATACCAAAAAACATTTGAT 1563
QY 241 CTATATACATAGACATGAATATATTTCTGTGTGTTGTCATATATACCTCAAC 300
Db 1562 CTATATACATAGACATGAATATATTTCTGTGTGTTGTCATATATACCTCAAC 1503
QY 301 ACTATTATTAATGCAATCTATATTTAGGTATAGAAGTTGATGATACCTTTCTAC 360
Db 1502 ACTATTATTAATGCAATCTATATTTAGGTATAGAAGTTGATGATACCTTTCTAC 1443
QY 361 TTGCCATGGCATTTAAACAAGCAAGGCTGAGACTCAGCAACCACTTGTTGTTCAATGCAATG 420
Db 1442 TTGCCATGGCATTTAAACAAGCAAGGCTGAGACTCAGCAACCACTTGTTGTTCAATGCAATG 1383
QY 421 CAGGCTAGTAGTAAGTTTGGTTGCTAGGAAAAGGCTCTTATCTCACCCCTCCTTAA 480
Db 1382 CAGGCTAGTAGTAAGTTTGGTTGCTAGGAAAAGGCTCTTATCTCACCCCTCCTTAA 1323
QY 481 ACTAAAGGTTCTTTCAGGCT 500
Db 1322 ACTAAAGGTTCTTTCAGGCT 1303

RESULT 10
AAH46959
ID AAH46959 standard; cDNA; 2736 BP.
XX
AC AAH46959;
XX
DT 25-SEP-2001 (first entry)
XX
DE Human secreted protein encoding cDNA (clone Id HGCNC48).
XX
KW Secreted protein; immunosuppressive; antiarthritic; antirheumatic;
KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
KW nootropic; neuroprotective; antibacterial; virucide; fungicide; human;
KW ophthalmological; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN WO200155430-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01431.
XX
PP 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.
PR 12-SEP-2000; 2000US-0231968.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Komatsoulis GA, Baker KP, Birse CE, Soppet DR, Olsen HS;
PI Moore PA, Wei P, Ebner R, Duan DR, Shi Y, Choi GH, Fiscella M;
PI Ni J, Ruben SM, Barash SC;
XX
XX WPI; 2001-476220/51.
DR P-PSDB; AAB85549.
XX
XX 17 isolated nucleic acid molecules encoding human secreted proteins,
PT used to preventing, treating or ameliorating a medical condition
XX
XX Claim 1; Page 433; 482pp; English.
XX
XX The invention provides novel human secreted proteins and polynucleotides
CC encoding them. The secreted proteins can be expressed by standard
CC recombinant methodology. The secreted proteins and polynucleotides are
CC used to prevent, treat or ameliorate a medical condition in e.g. humans,
CC mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They can
CC also be used in diagnosing a pathological condition. The antibodies to
CC the proteins can also be used in alleviating symptoms associated with the
CC disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme
CC linked immunosorbent assays (ELISA). Disorders which are diagnosed or
CC treated include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi and
CC ocular disorders e.g. corneal infection. The polypeptides can also be
CC used to aid wound healing and epithelial cell proliferation, to prevent
CC skin aging due to sunburn, to maintain organs before transplantation, for
CC supporting cell culture of primary tissues, to regenerate tissues and in
CC chemotaxis. The polypeptides can also be used as a food additive or
CC preservative to increase or decrease storage capabilities. The present
XX sequence represents a human secreted protein encoding cDNA.
SQ Sequence 2726 BP; 881 A; 465 C; 489 G; 886 T; 5 other;

Query Match 99.5%; Score 497.6; DB 22; Length 2726;
Best Local Similarity 99.4%; Pred. NO. 1e-95;
Matches 497; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 GTATATGTAAGAAAGCCCTCATCTTTTGGTTTCAAAAATTTAAATATACAGATGCTTCTTAAGAGA 60
DB |||||||
QY 61 GCAAGATTCAAAATTTGTTTGGTTTCAAAAATTTAAATATACAGATGCTTCTTAAGAGA 120
DB |||||||
QY 121 CTAAGACATGTTTCATATATTTGACCAATCCCTTATTTTGGCAAGAGATTTTAAGAGTCT 180
DB |||||||
QY 2122 CTAAGACATGTTTCATATATTTGACCAATCCCTTATTTTGGCAAGAGATTTTAAGAGTCT 2181
DB |||||||
QY 181 AACTCAACATATGTAAGCTCTGCTGCTACCTGGTTATATATACCAAAAAACATTTGAT 240
DB |||||||
QY 2182 AACTCAACATATGTAAGCTCTGCTGCTACCTGGTTATATATACCAAAAAACATTTGAT 2241
DB |||||||
QY 241 CTATATACACATAGACATGAATATATTTCTGTTGTTTGGTATATATACCTCAAC 300
DB |||||||
QY 2242 CTATATACACATAGACATGAATATATTTCTGTTGTTTGGTATATATACCTCAAC 2301
DB |||||||
QY 301 ACTATATTAATGCAATCCCTATATTTCTAGGTATAGAGTTGATATACCTTTCTAC 360
DB |||||||
QY 2302 ACTATATTAATGCAATCCCTATATTTCTAGGTATAGAGTTGATATACCTTTCTAC 2361
DB |||||||
QY 361 TTGCCATGGCATTAACAAGCAAGGCTGAGACTCAGCAACCACTTGTTTCATTGCAATTG 420
DB |||||||
QY 2362 TTGCCATGGCATTAACAAGCAAGGCTGAGACTCAGCAACCACTTGTTTCATTGCAATTG 2421
DB |||||||

QY 421 CAGCTAGTAGTAAGTTTGGTTTGGTGTAGGAAAAGGTCCTCTTATCTACCCCTCTTAA 480
DB |||||||
QY 2422 CAGCTAGTAGTAAGTTTGGTTTGGTGTAGGAAAAGGTCCTCTTATCTACCCCTCTTAA 2481
DB |||||||
QY 481 ACTAAAGGTTCTTTTCAGGCT 500
DB |||||||
QY 2482 ACTAATGGTTCTTTTCAGGCT 2501
DB |||||||
ID AAF24196 standard; DNA; 3853 BP.
XX
XX AAF24196;
XX
XX 02-APR-2001 (first entry)
XX Human VEGF-G cDNA.
XX Vascular endothelial growth factor; VEGF; cancer; cell;
KW angiogenesis; ss.
XX
XX Homo sapiens.
XX
XX W020000087/8-A2
XX
XX 01 JAN-2001.
XX
XX 29-JUN-2000; 2000WO-US18085.
XX
XX 30-JUN-1999; 99US-0343671.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Gearing DP;
XX
XX WPI; 2001-050129/06.
XX
XX New vascular endothelial growth factor family member used for diagnosis
PT and treatment of deregulated cell growth e.g. cancer, disorders
PT involving aberrant angiogenesis e.g. psoriasis, and chronic
PT inflammatory diseases
XX
XX Claim 1; Fig 1; 142pp; English.
XX
XX The present invention relates to a vascular endothelial growth factor
CC (VEGF) family member, VEGF-G. VEGF-G protein and nucleic acid
CC molecules are used as modulating agents or as targets for
CC developing modulating agents to regulate a variety of cellular
CC processes e.g. cell proliferation, differentiation, migration and
CC wound repair. VEGF-G modulators, i.e. VEGF-G protein, peptide,
CC peptidomimetic or nucleic acid are used to treat a subject with
CC aberrant VEGF-G protein or nucleic acid expression or activity
CC e.g. deregulated cell growth, such as cancer, hyperproliferic bone
CC disorders, disorders involving aberrant angiogenesis e.g. psoriasis,
CC and chronic inflammatory diseases e.g. rheumatoid arthritis. VEGF-G
CC gene expression is inhibited through the administration of antisense
CC molecules or ribozymes and by targeting the regulatory region of VEGF-G
CC to prevent transcription of the gene in target cells.
SQ Sequence 3853 BP; 1194 A; 771 C; 757 G; 1131 T; 0 other;

Query Match 99.5%; Score 484.8; DB 22; Length 3853;
Best Local Similarity 99.4%; Pred. NO. 5.2e-93;
Matches 498; Conservative 10; Mismatches 2; Indels 2; Gaps 1;
QY 1 GTATATGTAAGAAAGCCCTCATCTTTTGGTTTCAAAAATTTAAATATACAGATGCTTCTTAAGAGA 60
DB |||||||
QY 1851 GTATATGTAAGAAAGCCCTCATCTTTTGGTTTCAAAAATTTAAATATACAGATGCTTCTTAAGAGA 1792
DB |||||||
QY 61 GCAAGATTCAAAATTTGTTTGGTTTCAAAAATTTAAATATACAGATGCTTCTTAAGATT 120
DB |||||||
QY 1791 GCAAGATTCAAAATTTGTTTGGTTTCAAAAATTTAAATATACAGATGCTTCTTAAGATT 1732
DB |||||||

QY 121 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTGGCAAGGATTTTAAAGAGTCT 180
D 1731 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTGGCAAGGATTTTAAAGAGTCT 1672
QY 181 AACTCAACATATGTAAGCTGCTGGTGACCTGATATATATAC--AAAAAACATTTG 238
D 1671 AACTCAACATATGTAAGCTGCTGGTGACCTGATATATATATACAAAAAACATTTG 1612
QY 239 ATCTATATACATAGACATGAATATTTCTGTGCTGTTGTCGCATATATACCTCAA 298
D 1511 ATCTATATACATAGACATGAATATTTCTGTGCTGTTGTCGCATATATACCTCAA 1552
QY 299 ACCTATTATTAATGCAATCCCTATATTTCTAGGTATAGAGTTGATGATATACCTTTCT 358
D 1551 ACCTATTATTAATGCAATCCCTATATTTCTAGGTATAGAGTTGATGATATACCTTTCT 1492
QY 359 ACTTGCCATGGCATTACAAAGCAAGCTGAGACTCAGCAACCACTGTTGTTTCAATGTCAT 418
D 1491 ACTTGCCATGGCATTACAAAGCAAGCTGAGACTCAGCAACCACTGTTGTTTCAATGTCAT 1432
QY 419 TGCAGGCTACTAGTAAAGTTTGGTGGTGGTAGGAAAGGCTCTTATCTCACCCCTCT 478
D 1431 TGCAGGCTACTAGTAAAGTTTGGTGGTGGTAGGAAAGGCTCTTATCTCACCCCTCT 1372
QY 479 AAACATAAGGTTCTTTCAGGCT 500
D 1371 AAACATAAGGTTCTTTCAGGCT 1350

RESULT 12

AAC81555/c

ID AAC81555 standard; cDNA; 1882 BP.

AC AAC81555;

XX 09-MAR-2001 (first entry)

XX Human growth factor homologue zveg4 cDNA, SEQ ID NO:1.

XX Human; zveg4; growth factor homologue; VEGF/PDGF family;
XX CUB domain; PDGF-like activity; mitogenic; osteogenic;
XX neovascularisation; tissue repair; proliferation; differentiation;
XX liver damage; neurodegenerative; Alzheimer's disease; multiple sclerosis;
XX periodontal disease; bone fracture; wound healing; vulnary; ischaemia;
XX immunomodulation; hepatic; chromosome 11q22.3-23.1; ss.

XX Homo sapiens.

OS W020006736-A1.

XX 09-NOV-2000.

XX 03-MAY-2000; 2000WO-US40047.

XX 03-MAY-1999; 99US-0304216.

XX 10-NOV-1999; 99US-0164463.

XX 04-FEB-2000; 2000US-0180169.

XX (ZYMO) ZYMOGENETICS INC.

XX Gilbert T, Hart CE, Sheppard PO, Gilbertson DG;

XX WPI; 2000-687541/67.

XX P-PSDB; AAB48653.

XX Growth factor homologs and the nucleic acids that encode them, useful

XX e.g. for treating liver damage, ischaemia, multiple sclerosis and

XX Alzheimer's disease

XX Claim 35; Page 106-110; 143pp; English.

XX The invention relates to the human growth factor homologue zveg4

XX

CC (AAB48653), and nucleic acids encoding it (AAC81555). Zveg4 is a member
CC of the PDGF (platelet-derived growth factor)/VEGF (vascular endothelial
CC growth factor) family. Zveg4 has a growth factor domain (AAB48654)
CC characterised by a PDGF cysteine knot structure, and a CUB domain
CC (AAB48655) which has a beta barrel structure. Zveg4 has PDGF-like
CC activity, having mitogenic activity on fibroblasts, vascular smooth
CC muscle cells and pericytes, and has also been shown to stimulate bone
CC growth. The invention also relates to fusion proteins comprising human
CC zveg4 or fragments thereof, particularly human zveg4/human zveg3
CC fusions; expression constructs and host cells comprising human zveg4
CC nucleic acids; the recombinant expression of human zveg4; an antibody
CC which binds to human zveg4 or a fragment thereof; a method of activating
CC a cell-surface PDGF receptor using a zveg4-derived polypeptide; a
CC method of modulating the proliferation, differentiation, migration or
CC metabolism of bone cells, comprising exposing bone cells to
CC zveg4-derived polypeptides; and a method of detecting a genetic
CC abnormality in the zveg4 gene of a patient. Zveg4 proteins and derived
CC fragments may be used to stimulate tissue development or repair, or
CC cellular differentiation or proliferation. They are particularly used for
CC the treatment or repair of liver damage, and may also be used to
CC modulate neurite growth (e.g., in the treatment of Alzheimer's disease or
CC multiple sclerosis). Due to their osteogenic activity, they may be used
CC in the treatment of periodontal disease and fractures. They may also be
CC used to enhance expansion and mobilisation of haematopoietic stem cells
CC and endothelial precursor stem cells, which may be useful in the
CC treatment of ischaemia, in wound healing, and in the modulation of the
CC immune system. The present sequence represents cDNA encoding human
CC zveg4.
XX

SQ Sequence 1882 BP; 566 A; 407 C; 430 G; 479 T; 0 other;

Query Match 96.6%; Score 482.8; DB 21; Length 1882;

Best Local Similarity 98.8%; Pred. No. 1.3e-92;

Matches 498; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

QY 1 GTATATGTAAGAAAGCCTCATCTTTTGATTTTAAATATACAAAGATGTTCTTTAAGAGA 60
D 1866 GTATATGTAAGAAAGCCTCATCTTTTGATTTTAAATATACAAAGATGTTCTTTAAGAGA 1807
QY 61 GCAAGATTCAAAATTTGTTTGTGTTTCAAAATTTAAATAAATTTATCTCCCTAAATTTT 120
D 1806 GCAAGATTCAAAATTTGTTTGTGTTTCAAAATTTAAATAAATTTATCTCCCTAAATTTT 1747
QY 121 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTGGCAAGGATTTTAAAGAGTCT 180
D 1746 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTGGCAAGGATTTTAAAGAGTCT 1687
QY 181 AACTCAACATATGTAAGCTGCTGGTCTACCTGGTTATATATACCC---AAAAAACATTT 236
D 1686 AACTCAACATATGTAAGCTGCTGGTCTACCTGGTTATATATACCCAAAAAACATTT 1627
QY 237 TGATCTATATACATAGACATGAATATATTTCTGTGTTGTTGTCATATATAACCTC 296
D 1626 TGATCTATATACATAGACATGAATATATTTCTGTGTTGTTGTCATATATAACCTC 1567
QY 297 AAACACTATTTAAATGCAATCCTATATTTCTTAGGTATAGAAAGTTGATGATATACCTTT 356
D 1566 AAACACTATTTAAATGCAATCCTATATTTCTTAGGTATAGAAAGTTGATGATATACCTTT 1507
QY 357 CTACTTGGCCATGCAATTAACAAGCAAGGCTGAGACTCAGCAACCACTTGTGTTTCATTCG 416
D 1506 CTACTTGGCCATGCAATTAACAAGCAAGGCTGAGACTCAGCAACCACTTGTGTTTCATTCG 1447
QY 417 ATTGCAGGCTAGTAGTAAGTTTGGTCTGGTAGGAAAGGGTCTCTTATCTCACCTCC 476
D 1446 ATTGCAGGCTAGTAGTAAGTTTGGTCTGGTAGGAAAGGGTCTCTTATCTCACCTCC 1387
QY 477 TTAACATAAAGGTTCTTTTCAGGCT 500
D 1386 TTAACATAAAGGTTCTTTTCAGGCT 1363

RESULT 13

PR	04-FEB-2000; 2000US-180169P.
XX	31-MAR-2000; 2000US-054022A.
XX	(ZYMO) ZYMOGENETICS INC.
XX	Gilbertson DG, Hart CE;
PI	
XX	WPI; 2001-611088/70.
DR	P-PSDB; AAG65601.
XX	
PT	Use of zvegfa polypeptide for promoting bone, ligament or cartilage
PT	growth in mammal at site of fracture, implant, and bone graft, and for
PT	promoting growth or differentiation of osteoblasts, chondrocytes in
PT	culture
XX	
PS	Example 1.; Page 44-47; 57pp; English.
XX	
CC	The invention relates to the use of zvegfa polypeptide for promoting
CC	bone, ligament or cartilage growth in a mammal, and for promoting
CC	proliferation or differentiation of osteoblasts, osteoclasts,
CC	chondrocytes or bone marrow stem cells in culture. For promoting
CC	cartilage growth, chondrocytes are cultured ex vivo in presence of the
CC	zvegfa polypeptide and then placed into mammal where cartilage is to be
CC	grown. Zvegfa polypeptide is useful for promoting growth of bone,
CC	ligament or cartilage in a mammal at a site of bony defect such as
CC	fracture, bone graft, implant or periodontal pocket, in humans and non-
CC	human animals such as domestic animals including livestock and companion
CC	animals. Zvegfa is used for promoting growth of bone, ligament, or
CC	cartilage in conditions of bone defects following therapeutic treatments
CC	of bone cancers or other conditions characterized by increased bone loss,
CC	or decreased bone formation, or elevation of peak bone mass in pre-
CC	menopausal woman. It is also useful for healing bone following radiation
CC	-induced osteonecrosis, repairing bone defects arising from surgery, and
CC	promotion of bone healing in plastic surgery, increasing bone formation
CC	during distraction osteogenesis, treating bone injuries including repair
CC	of cartilage and ligament and treatment of osteoporosis. The present
CC	sequence represents a human zvegfa polypeptide encoding cDNA.
XX	
SQ	Sequence 1882 BP; 566 A; 407 C; 430 G; 479 T; 0 other;
Query Match 96.68; Score 482.8; DB 22; Length 1882;	
Best Local Similarity 98.8%; Pred. No. 1.3e-92;	
Matches 498; Conservative 0; Mismatches 2; Indels 4; Gaps 1;	
Qy	1 GTATATGTAAGAAGCCTCATCTTTTGATTGTTTTAAATACAAAGATGCTTCTTTAAGAGA 60
Dd	
Dd	1866 GTATATGTAGAAGCCTCATCTTTTGATTGTTTTAAATACAAAGATGCTTCTTTAAGAGA 1807
Qy	61 GCAAGATTCAAATTTGTTTGTGGTTCACAAATTTAAAAATAAATTTATCTCCCTAAATTTT 120
Dd	
Dd	1806 GCAAGATTCAAATTTGTTTGTGGTTCACAAATTTAAAAATAAATTTATCTCCCTAAATTTT 1747
Qy	121 CTAAGACATGTTTCATATATTTTGACCATCCCTTATTTTGGCAAGAGATTTTAAAGAGTCT 180
Dd	
Dd	1746 CTAAGACATGTTTCATATATTTTGACCATCCCTTATTTTGGCAAGAGATTTTAAAGAGTCT 1687
Qy	181 AACTCAAACATATGTAAGCTCTGGTGACCTGGTTATATATAC ---AAAAAAAAACATT 236
Dd	
Dd	1686 AACTCAAACATATGTAAGCTCTGGTGACCTGGTTATATATACCAAAAATAAACAATT 1627
Qy	237 TGATCTATACATACATAGACATGAATATATTTCTGTGTGTTTGTGCATATATAACCTC 296
Dd	
Dd	1626 TGATCTATACATACATAGACATGAATATATTTCTGTGTGTTTGTGCATATATAACCTC 1567
Qy	297 AAACACATTTATTAATGCAATCCCTATATTCCTTAGGTTATAGAGTTGATGATATACCTTT 356
Dd	
Dd	1566 AAACACATTTATTAATGCAATCCCTATATTCCTTAGGTTATAGAGTTGATGATATACCTTT 1507
Qy	357 CTACTTGCCATGCCATTAAACAAGCAAGCGTGAGACTCAGCACCACTTGTTTCATTGTC 416
Dd	
Dd	1506 CTACTTGCCATGCCATTAAACAAGCAAGCGTGAGACTCAGCACCACTTGTTTCATTGTC 1447
Qy	417 ATTGCGAGGCTAGTAGTAAGTTTGGTGTGCTAGGAAAAGGGTCTCTTTATCTCACCCCTCC 476

Db	1446	ATTGACGGCTAGTAGTAAGTTTGGTCTGCTAGTAAGGGTCTCTTATCTCACCGTCC	1387
Qy	477	TTAACTAAAGGTTCTTTTCAGGCT	500
Db	1386	TTAACTAAAGGTTCTTTTCAGGCT	1363
RESULT 15			
ABQ73239/c			
ID	ABQ73239	standard; cDNA; 1882 BP.	
XX	AC	ABQ73239;	
XX	DT	30-SEP-2002 (first entry)	
XX	XX	Human zvegfg4 encoding cDNA SEQ ID NO:1.	
XX	XX	Human; zvegfg4; cell proliferation; extracellular matrix production; fibroproliferative disorder; PDGF-D; platelet derived growth factor; PDGF; vascular endothelial growth factor; VEGF; cytotstatic; nephrotropic; hepatotropic; antiinflammatory; osteopathic; antiarthritic; metastasis; prostate tumour; prostate cancer; glomerulonephritis; lupus nephritis; diabetic glomerulosclerosis; renal arteriosclerosis; nephrotic syndrome; chronic active hepatitis; cirrhosis; osteopetrosis; osteosclerosis; hyperostosis; osteoarthritis; gene; ss.	
XX	OS	Homo sapiens.	
XX	XX		
XX	PH	Key	Location/Qualifiers
XX	FT	CDS	226..1338
XX	FT		/*tag= a
XX	FT		/product= "zvegfg4"
XX	XX		
XX	PN	US2002064832-A1.	
XX	XX		
XX	PD	30-MAY-2002.	
XX	XX		
XX	PF	14-MAR-2001; 2001US-0808972.	
XX	PR	03-MAY-1999; 99US-132250P.	
XX	PR	10-NOV-1999; 99US-164463P.	
XX	PR	04-FEB-2000; 2000US-180169P.	
XX	PR	26-SEP-2000; 2000US-235293P.	
XX	PR	03-MAY-2000; 2000US-056459S.	
XX	XX		
XX	PA	(HART/) HART C E.	
XX	PA	(TOPO/) TOPOUZIS S.	
XX	PA	(GILB/) GILBERTSON D G.	
XX	XX		
XX	PI	Hart CE, Topouzis S, Gilbertson DG;	
XX	XX		
XX	DR	WPI: 2002-573696/61.	
XX	DR	P-FSDB: ABP51640.	
XX	XX		
XX	PT	Reducing proliferation or extracellular matrix production by a cell in a mammal, useful for treating fibroproliferative disorders of bone, liver and kidney, comprises administering a zvegfg4 antagonist	
XX	XX	Example 1; Page 18-19; 34pp; English.	
XX	XX	The present invention describes a method for reducing proliferation of or extracellular matrix production by a cell in a mammal. The method comprises administering to the mammal a composition comprising a therapeutically effective amount of a zvegfg4 antagonist chosen from anti-zvegfg4 antibodies, inhibitory polynucleotides, inhibitors of zvegfg4 activation, and mitogenically inactive, receptor-binding variants of zvegfg4. Zvegfg4 (also called PDGF-B) is a multi-domain protein that is structurally related to platelet derived growth factor (PDGF) and vascular endothelial growth factors (VEGF). Zvegfg4 has cytostatic, nephrotropic, hepatotropic, antiinflammatory, osteopathic and antiarthritic activities. The method is useful for reducing proliferation of mesangial, epithelial, endothelial, smooth muscle, fibroblast,	

CC osteoblast, osteoclast, neuronal, stromal, stellate or interstitial cells
CC in a mammal, in particular proliferation of prostate tumour cells, and
CC for reducing extracellular matrix production by a cell in a mammal
CC suffering from a fibroproliferative disorder of kidney, bone or liver.
CC In particular it is useful for reducing stellate cell activation. The
CC method is useful for reducing metastasis of prostate cancer cells to
CC bone in a mammal and for treating a fibroproliferative disorder of
CC kidney, liver or bone in a mammal. Fibroproliferative disorders of the
CC kidney include, glomerulonephritis, diabetic glomerulosclerosis, lupus
CC nephritis, renal arteriosclerosis and nephrotic syndrome, disorders of
CC the liver include chronic active hepatitis and many other types of
CC cirrhosis, and disorders of the bone include osteopetrosis, hyperostosis,
CC osteosclerosis, osteoarthritis, and ectopic bone formation in metastatic
CC prostate cancer. The present sequence encodes human zvegfa4, which is
CC used in an example from the present invention.

XX
SQ Sequence 1882 BP; 566 A; 407 C; 430 G; 479 T; 0 other;

Query Match 96.6%; Score 482.8; DB 24; Length 1882;

Best Local Similarity 98.8%; Pred. No. 1.3e-92;

Matches 498; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

QY	1	GTATATGTAAGAAACCTCATCTTTTGATTTTATATACAGATGCTTCTTTAAGAGA	60
DB	1866	GTATATGTAAGAAACCTCATCTTTTGATTTTATATACAGATGCTTCTTTAAGAGA	1807
QY	61	GCAAGATTCAAAATGTTTTGTTTCAAAATTTAAAAATAAATTTATCTCTAAATTTT	120
DB	1806	GCAAGATTCAAAATGTTTTGTTTCAAAATTTAAAAATAAATTTATCTCTAAATTTT	1747
QY	121	CTAAGACATGTTTCATATATTGACCATCCCTTATTTGGCAAGAGATTTTAAGAGTCT	180
DB	1746	CTAAGACATGTTTCATATATTGACCATCCCTTATTTGGCAAGAGATTTTAAGAGTCT	1687
QY	181	AACTCAACATATGTAAGCTCTGGTGTAACCTGGTTATATATACC----	236
DB	1686	AACTCAACATATGTAAGCTCTGGTGTAACCTGGTTATATATACC-----	1627
QY	237	TGATCTATATACATAGACATGAATATATTCTGTGTGTGTGTGCATATATAACCTC	296
DB	1626	TGATCTATATACATAGACATGAATATATTCTGTGTGTGTGTGTGCATATATAACCTC	1567
QY	297	AAACACTATTATAATGCAATCCCTATATTCTTAGGTATAGAAGTTGATGATACCTTT	356
DB	1566	AAACACTATTATAATGCAATCCCTATATTCTTAGGTATAGAAGTTGATGATACCTTT	1507
QY	357	CTACTTGCATGGCATTAAACCAAGCAAGCTGAGACTCAGCAACCACCTGTGTCATTGC	416
DB	1506	CTACTTGCATGGCATTAAACCAAGCAAGCTGAGACTCAGCAACCACCTGTGTCATTGC	1447
QY	417	ATTGAGGCTAGTAGTAAGTTGGTTGCTGGTAGGAAAGGTCCTTTATCTCACCCCTCC	476
DB	1446	ATTGAGGCTAGTAGTAAGTTGGTTGCTGGTAGGAAAGGTCCTTTATCTCACCCCTCC	1387
QY	477	TTAAACTAAAGGTTCTTTTCAGGCT	500
DB	1386	TTAAACTAAAGGTTCTTTTCAGGCT	1363

Search completed: July 4, 2003, 15:25:01

Job time : 168.445 secs

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
C 1	482.8	96.6	1882	4	US-09-540-224-1	Sequence 1, Appli	
C 2	482.8	96.6	1910	4	US-09-457-066-36	Sequence 36, Appli	
C 3	74.2	14.8	1472	4	US-09-540-224-3	Sequence 3, Appli	
C 4	61.4	12.3	256	4	US-09-222-575-13	Sequence 13, Appli	
C 5	49.4	9.9	19124	2	US-08-487-826B-13	Sequence 13, Appli	
C 6	42	8.4	19124	2	US-08-487-826B-13	Sequence 13, Appli	
C 7	40.8	8.2	1422	1	US-08-319-704-5	Sequence 5, Appli	
C 8	40.2	8.0	162450	4	US-09-345-882-1	Sequence 1, Appli	
C 9	40	8.0	2435	4	US-09-306-593-1	Sequence 1, Appli	
C 10	39.8	8.0	168575	4	US-09-426-290-1	Sequence 1, Appli	
C 11	39.6	7.9	401	4	US-09-643-597-262	Sequence 262, App	
C 12	39.6	7.9	5433	3	US-08-929-329-1	Sequence 1, Appli	
C 13	39.4	7.9	1368	3	US-08-874-563-5	Sequence 5, Appli	
C 14	39.4	7.9	1368	3	US-08-577-483-14	Sequence 14, Appli	
C 15	39.2	7.8	6152	4	US-08-973-462-1	Sequence 1, Appli	
C 16	39	7.8	6350	4	US-09-647-826-1	Sequence 1, Appli	
C 17	38.8	7.8	162450	4	US-09-345-882-1	Sequence 1, Appli	
C 18	38.6	7.7	1016	1	US-08-109-391A-3	Sequence 3, Appli	
C 19	38.6	7.7	1016	1	US-08-459-019A-3	Sequence 3, Appli	
C 20	38.6	7.7	1016	2	US-08-460-428A-3	Sequence 3, Appli	
C 21	38.6	7.7	1016	3	US-08-458-860A-3	Sequence 3, Appli	
C 22	38.6	7.7	2659	3	US-08-749-522-1	Sequence 1, Appli	
C 23	38.6	7.7	2755	3	US-08-749-522-2	Sequence 2, Appli	
C 24	38.6	7.7	8920	2	US-08-446-855A-1	Sequence 1, Appli	
C 25	38.6	7.7	8920	2	US-08-446-855A-1	Sequence 1, Appli	
C 26	38.6	7.7	8920	4	US-09-150-741-1	Sequence 1, Appli	
C 27	38.6	7.7	8920	4	US-09-150-741-1	Sequence 1, Appli	

APPLICANT: Xu, Jiangchun
TITLE OF INVENTION: Compositions for the Treatment and Diagnosis of Breast Cancer

QY 192 ATGTAAGCTCTCGTGTACCTGGTTATATATACCAAAAAACATTTGATCTATATACACA 251

121 CTAAGACATGTTTCATATATTGACCATCCCTTA
QY

Qy 121 CTAAGACATGTTTCATATATTGACCATCCCTATTTTGCAAGAGSATTTTAAAGAGTCT 18


```
NAME/KEY: allele
LOCATION: 108106 : polymorphic base insertion of A
OTHER INFORMATION: 5-135-155 : polymorphic fragment 5-128-60 SEQ ID53
FEATURE:
NAME/KEY: allele
LOCATION: 108149 : polymorphic base insertion of GTTT
OTHER INFORMATION: 5-135-198 : polymorphic fragment 99-1442-224 SEQ ID50
FEATURE:
NAME/KEY: allele
LOCATION: 108308 : polymorphic base A or G
OTHER INFORMATION: 5-135-357 : polymorphic fragment 99-1442-224 SEQ ID71
FEATURE:
NAME/KEY: allele
LOCATION: 108471 : polymorphic base C or T
OTHER INFORMATION: 5-136-174 : polymorphic fragment 5-129-144 SEQ ID33
FEATURE:
NAME/KEY: allele
LOCATION: 134134 : polymorphic base C or T
OTHER INFORMATION: 5-140-120 : polymorphic fragment 5-129-144 SEQ ID54
FEATURE:
NAME/KEY: allele
LOCATION: 134362 : polymorphic base insertion of A
OTHER INFORMATION: 5-140-348 : polymorphic fragment 5-130-257 SEQ ID34
FEATURE:
NAME/KEY: allele
LOCATION: 134374 : polymorphic base insertion of CA
OTHER INFORMATION: 5-140-361 : polymorphic fragment 5-130-257 SEQ ID55
FEATURE:
NAME/KEY: allele
LOCATION: 146328 : polymorphic base A or G
OTHER INFORMATION: 5-143-84 : polymorphic fragment 5-130-276 SEQ ID35
FEATURE:
NAME/KEY: allele
LOCATION: 146345 : polymorphic base A or C
OTHER INFORMATION: 5-143-101 : polymorphic fragment 5-130-276 SEQ ID56
FEATURE:
NAME/KEY: allele
LOCATION: 150329 : polymorphic base A or G
OTHER INFORMATION: 5-145-24 : polymorphic fragment 5-131-395 SEQ ID36
FEATURE:
NAME/KEY: allele
LOCATION: 160031 : polymorphic base G or T
OTHER INFORMATION: 5-148-352 : polymorphic fragment 5-131-395 SEQ ID57
FEATURE:
NAME/KEY: allele
LOCATION: 72771..72817 : polymorphic fragment 5-124-273 SEQ ID30
OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID51
FEATURE:
NAME/KEY: allele
LOCATION: 72771..72817 : polymorphic fragment 5-124-273 SEQ ID51
OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID51
FEATURE:
NAME/KEY: allele
LOCATION: 88050..88096 : polymorphic fragment 5-127-261 SEQ ID31
OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID52
FEATURE:
NAME/KEY: allele
LOCATION: 88050..88096 : polymorphic fragment 5-127-261 SEQ ID52
OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID52
FEATURE:
NAME/KEY: allele
LOCATION: 90819..90865 : complement polymorphic fragment 99-1437-325 SEQ ID49
OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID49
FEATURE:
NAME/KEY: allele
LOCATION: 90819..90865 : complement polymorphic fragment 99-1437-325 SEQ ID70
OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID70
FEATURE:
NAME/KEY: allele
LOCATION: 93690..93736 : polymorphic fragment 5-128-60 SEQ ID32
OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID32
FEATURE:
NAME/KEY: allele
LOCATION: 93690..93736 : polymorphic fragment 5-128-60 SEQ ID53
OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID53
FEATURE:
NAME/KEY: allele
LOCATION: 97099..97145 : polymorphic fragment 99-1442-224 SEQ ID50
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID50
FEATURE:
NAME/KEY: allele
LOCATION: 97099..97145 : polymorphic fragment 99-1442-224 SEQ ID71
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID71
FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177 : polymorphic fragment 5-129-144 SEQ ID33
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID33
FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177 : polymorphic fragment 5-129-144 SEQ ID54
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID54
FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121 : polymorphic fragment 5-130-257 SEQ ID34
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID34
FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121 : polymorphic fragment 5-130-257 SEQ ID55
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID55
FEATURE:
NAME/KEY: allele
LOCATION: 99094..99140 : polymorphic fragment 5-130-276 SEQ ID35
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID35
FEATURE:
NAME/KEY: allele
LOCATION: 99094..99140 : polymorphic fragment 5-130-276 SEQ ID56
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID56
FEATURE:
NAME/KEY: allele
LOCATION: 103783..103828 : polymorphic fragment 5-131-395 SEQ ID36
OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID36
FEATURE:
NAME/KEY: allele
LOCATION: 103783..103828 : polymorphic fragment 5-131-395 SEQ ID57
OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID57
FEATURE:
NAME/KEY: allele
LOCATION: 106918..106966 : polymorphic fragment 5-133-375 SEQ ID37
OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID37
FEATURE:
NAME/KEY: allele
LOCATION: 106918..106966 : polymorphic fragment 5-133-375 SEQ ID58
OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID58
FEATURE:
NAME/KEY: allele
LOCATION: 108084..108130 : polymorphic fragment 5-135-155 SEQ ID38
OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID38
FEATURE:
NAME/KEY: allele
LOCATION: 108084..108130 : polymorphic fragment 5-135-155 SEQ ID59
OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID59
FEATURE:
NAME/KEY: allele
LOCATION: 108127..108177 : polymorphic fragment 5-135-198 SEQ ID39
OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID39
FEATURE:
NAME/KEY: allele
LOCATION: 108127..108177 : polymorphic fragment 5-135-198 SEQ ID60
OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID60
FEATURE:
```

Query Match 8.0%; Score 40.2; DB 4; Length 162450;
Best Local Similarity 56.4%; Pred. No. 0.97; Mismatches 0; Gaps 0;
Matches 75; Conservative 0

```

; CURRENT FILING DATE: 1999-10-25
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 168575
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (21181)...(21403)
; NAME/KEY: CDS
; LOCATION: (95252)...(95430)
; NAME/KEY: CDS
; LOCATION: (101753)...(101996)
; NAME/KEY: CDS
; LOCATION: (110324)...(110439)
; NAME/KEY: CDS
; LOCATION: (124058)...(124278)
; NAME/KEY: CDS
; LOCATION: (127009)...(127130)
; NAME/KEY: CDS
; LOCATION: (128910)...(129139)
; US-09-426-290-1

Query Match      8.0%; Score 39.8%; DB 4; Length 168575;
Best Local Similarity 46.3%; Pred. No. 1.2;
Matches 131; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

Qy    24   TTGTGATTTTAAATACAGAAGATCTTTCTTTTAAGAGAGACAAGATTCAAATAATTGTTTTGTG 83
Db    109148 TATTATATATATATATAAAGTTATATATATATATAATATATAATAAGTTATATATATATATATATATATATATATAT 109089

Qy    84   TTTCAAAATTTAAAAATAAATTTATCTCCTAAATTTTCTAAAGACATGTTTCATATATATT 143
Db    109088 TTAGATATATATAAATAAATATATATTAATAGATATTTATATGTAATATATATATAACA 109029

Qy    144   GACCATCCCTTTTTCGGCAAGATTTTAAAGAGTCCTAACCTCAACACATATGTAAGCTCTG 203
Db    109028 TATAATACATATATATATTTATATAATATATATAAATATATATAACTATAATACATATATATTTA 108969

Qy    204   GTGTACCTGGTTATATATACCAAAAAACATTTTGATCTATATACACATAGACATGAATA 263
Db    108968 TATATAATATATATATAACATATAATATATATATATATATATATATATATATATATATATGTCGGGATCTGAATT 108909

Qy    264   TATTTCTGTGTGTGTTTGTCATATATAAACCTCAAACACTATT 306
Db    108908 TTTCGTATGTATGTTTTTTTTTAATTTTATTCGGATAGGTTT 108866

RESULT 11
US-09-643-597-262
; Sequence 262, Application US/09643597
; Patent No. 6426072
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Fan, Liqun
; APPLICANT: Kalos, Michael D.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Hosken, Nancy
; APPLICANT: Fanger, Gary R.
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Henderson, Robert A.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.455C11
; CURRENT APPLICATION NUMBER: US/09/643,597
; CURRENT FILING DATE: 2000-08-21
; NUMBER OF SEQ ID NOS: 369
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 262

```

```

; LENGTH: 401
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(401)
; OTHER INFORMATION: n = A,T,C or G
US-09-643-597-262

Query Match
Best Local Similarity 7.9%; Score 39.6; DB 4; Length 401;
Matches 144; Conservative 0; Mismatches 177; Indels 0; Gaps 0;

QY 63 AAGATTCAAATTTGTTGTGTTTCAAAATTTAAATAATTTATCTCTCTAAATTTCT 122
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8 AACATTTCTAATATTTTGGCTTTTCATATATCAAGAGAGATTATGTGAACATATTTTAA 67
QY 123 AAAGACATGTTTCATATATTTCACCATCCCTTATTTTGGCAAGGATTTTAAGAGCTAA 182
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
68 ATACTGTAAGTGACATATAGTTATAAGATATATTTCTGTACAGTAGAGAAAGATTAT 127
QY 183 CTCAAACATATGTAAGCTCTGTGTACCTGGTTATATATACCAAAAAAACATTTGATCT 242
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
128 ACATGCAAGATATTGTACCATATACATTTTCATCTCGATCTCATAGAATAATCAAAA 187
QY 243 ATATACATACATGACATGAATATATTCTGTGTGTGTTGTGCATATATAACCTCAACAC 302
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
188 GAATAATGATACAGGTGAAATATGTTTACTTTCTCTAAATCAAGCTAGTTGTCACTC 247
QY 303 TATTATTAAATGCAATCCTATATTCTTAGGTATAGAAGTTGATGATATACCTTTCTACTT 362
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
248 AAAAATATGNTGCATAGTTTATTTTGAATTTAGGTTTGGGACTACTTTTTCCTCANT 307
QY 363 GCATGGCATTAACAAAGCAA 383
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
308 TCAATGAGAAATAAATCTA 328

RESULT 12
US-08-929-329-1
; Sequence 1, Application US/08929329
; Patent No. 6120770
; GENERAL INFORMATION:
; APPLICANT: Adams, John H
; APPLICANT: Dalton, John P
; APPLICANT: Kappe, Stefan
; TITLE OF INVENTION: Plasmodium Proteins Useful for Preparing
; TITLE OF INVENTION: Vaccine Compositions
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Barnes & Thornburg
; STREET: 11 S Meridian
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/929,329
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Breen, John P
; REGISTRATION NUMBER: 38,833
; REFERENCE/DOCKET NUMBER: 835910-28685
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 231-7745
; TELEFAX: (317) 231-7433
; INFORMATION FOR SEQ ID NO: 1:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 5433 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium yoelii
US-08-929-329-1

Query Match
Best Local Similarity 7.9%; Score 39.6; DB 3; Length 5433;
Matches 66; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 242 TATATACATACATGACATGAATATATTTCTGTGTGTGTTGTGCATATATAACCTCAACA 301
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5114 TATATATGATACAAACGTATATATTTATGTCTATTTCAAAATAAATATACATATTTCTAAA 5173
QY 302 CTATTATTAAATGCAATCCTATATTTCTTAGGTATAGAAGTTGATGATATA 351
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5174 GAATTATAAATATTATGCAATTTTCAATAAATAAATAAATTTTGTATTA 5223

RESULT 13
US-08-874-563-5/c
; Sequence 5, Application US/08874563
; Patent No. 6002071
; GENERAL INFORMATION:
; APPLICANT: CHAPPELL, JOSEPH
; APPLICANT: NEWMAN, JEFFREY D.
; APPLICANT: YIN, SHAOHUI
; TITLE OF INVENTION: TRANSCRIPTIONAL SILENCING
; TITLE OF INVENTION: ELEMENTS AND THEIR BINDING FACTORS
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Clark & Elbing LLP
; STREET: 176 Federal Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/874,563
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/020,087
; FILING DATE: June 13, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul T. Clark, Esq.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 07678/008001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-428-0200
; TELEFAX: 617-428-7045
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1368 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
US-08-874-563-5

Query Match
Best Local Similarity 7.9%; Score 39.4; DB 3; Length 1368;
Matches 44; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

```

Matches	151;	Conservative	0;	Mismatches	186;	Indels	0;	Gaps	0;
Qy	52	TTTAAAGAGACAAGATTCAA	AAATTTGTTTCTGTTTCAAAATTTAAAAATTTAAATTTATCTC	111					
Db	719	TGTACAAAGCAAGCTTATGCT	TTTGTCTCCACACAAAAAATGAAATTTTAA	660					
Qy	112	CTAAATTTCTAAGACATGTTTCATATATTTGACCATCCCTTAATTTGGCAAAGATTT	171						
Db	659	TATATTAATTTCAAGTTAGTTTTTGAATTTAAATTTCCCTTGGTATCGCTCTAGTTTTT	600						
Qy	172	TAAAGTCTTAACCTCAACACATATGTAAGCTCTGGTGACCTGGTTATATATACCAAAAAA	231						
Db	599	ATATTGTTTTTGATCACATAGGTATATGTATCTTTACGCCCTTTTATTTTTCAGAAAAA	540						
Qy	232	ACATTTGATCTATATACATACATACATGAATATATTTCTGTGCTGTTTGTGCATATATA	291						
Db	539	AAAGTTGATGGTGGCGGATGGATACACACACACACATACATTTATATATATA	480						
Qy	292	ACCTCAAACTATTTAAATGCAATCCCTATATTTCTTAGGTATAGAAGTTGATGATATA	351						
Db	479	TATATATATATATATATATATATATATATATATTTGTTTGTACACATAGGTATAATGTA	420						
Qy	352	CTTTTCTACPTGGCATGGCATTTAACAAAGCAAGGCTG	388						
Db	419	TCCTTACGCCCTTTTATTTTTTCAAGAAAAAAAAGTTG	383						

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999

```

Query Match 7.9%; Score 39.4; DB 3; Length 1368;
Best Local Similarity 44.8%; Pred. No. 0.54; .

Matches	151; Conservative	0; Mismatches	186; Indels	0; Gaps	0;
QY	52	TTTAAAGAGACAAAGATTCAAATTTGTTTTGTTTCAAAATTTAAAAATTAATTTATCTC	111		
Db	719	TGTTACAAAGCAAAAGTTATGCTTTGTTTCTCCACACAAAAAATCAAAATTTTAAA	660		
QY	112	CTAAATTTTCTAAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAGGATTT	171		
Db	659	TATATTAAATTTCAAGTTAGTTTTTGAATTTTAAATTTTCCCTTGGTATCGCTCTAGTTTTT	600		
QY	172	TRAGAGTCTAACTCAACACATATGTAAGCTCTGGGTACCTGGTTATATATACCAAAAAA	231		
Db	599	ATATCTTTTGATGACATAGGTATATATGTAATCTTTACGCCCTTTATTTTTTCAAGAAAA	540		
QY	232	ACATTTTGATCTATATACACATAGACATGAATATATTTCTGTGTGTTTGTGCATATATA	291		
Db	539	AAAGTTGCATGGTCGGCGCATCGGATACACACACACACATACATTTATATATATA	480		
QY	292	ACCTCAAAACACTATTATTAATCAATCCCTATATCTTACGGTATAGAAGTTGATGATATA	351		
Db	479	TATATATATATATATATATATATATATATATATTTGTTTGTAGCATAGGTATAATGTA	420		
QY	352	CCTTTCTACTTGGCATGGCAATTAACAAGCAAGGCTG	388		
Db	419	TCCTTTACGGCTTTTATTTTTTTTCAAGAAAAAAGTTG	383		

```

RESULT 15
US-08-973-462-1
; Sequence 1, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B
; CURRENT FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 6152
; TYPE: DNA
; ORGANISM: P. falciparum
US-08-973-462-1

```

Query Match	7.8%	Score 39.2;	DB 4;	Length 6152;
Best Local Similarity	56.1%	Pred. No. 0.83;		
Matches	74;	Conservative 0;	Mismatches 58;	Indels 0; Gaps 0;
QY	225	AAAAAACAATTGATCTATATACATACATACAGATGAATATATTTCTGTGTGTGTTCTGC	284	
Db	5690	ATATAAATAAATAAATAACATACATATATATATATATATATATCTTTTACAA	5749	
QY	285	ATATATAACCTCAAAACACATTTATTAATGCAATCCTATATTTCTTAGGTATAGAAGTTGA	344	
Db	5750	AATTTTAAAATTTTAAAATTTATATATATTAATTTATATATTTTCCATATATAATTTTA	5809	
QY	345	TGATATACCTTT	356	
Db	5810	TTTTTCAATTT	5821	

Search completed: July 4, 2003, 19:02:49
Job time : 36.9534 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 16:40:54 ; Search time 111.851 Seconds

(without alignments)
6978.003 Million cell updates/sec

Title: US-10-083-853b-2_COPY_1_500

Perfect score: 500

Sequence: i gtaatgtcaagaagcctca.....actaaagtttttcaggct 500

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1085931 seqs, 780495707 residues

Total number of hits satisfying chosen parameters: 2171862

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

Published_Applications_NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	500	100.0	1934	9	US-10-086-623-5	Sequence 5, Appli
C 2	500	100.0	1934	9	US-10-260-539-5	Sequence 5, Appli
C 3	500	100.0	2253	9	US-10-086-623-7	Sequence 7, Appli
C 4	500	100.0	2253	9	US-10-260-539-7	Sequence 7, Appli
C 5	500	100.0	3718	9	US-10-258-557-3	Sequence 3, Appli
C 6	500	100.0	3736	9	US-10-258-557-1	Sequence 1, Appli
C 7	500	100.0	2921	9	US-10-083-853-1	Sequence 1, Appli
C 8	499.6	99.9	4001	10	US-09-915-582-35	Sequence 35, Appli
C 9	498.4	99.7	3798	10	US-09-915-582-17	Sequence 17, Appli
C 10	497.6	99.6	2726	10	US-09-915-582-37	Sequence 37, Appli
C 11	482.8	96.6	1882	9	US-10-139-583-36	Sequence 36, Appli
C 12	482.8	96.6	1882	9	US-10-039-847A-1	Sequence 1, Appli
C 13	482.8	96.6	1882	9	US-10-226-559-1	Sequence 1, Appli
C 14	482.8	96.6	1882	9	US-10-274-638-1	Sequence 1, Appli
C 15	482.8	96.6	1882	10	US-09-808-972-1	Sequence 1, Appli
C 16	460	92.0	485	9	US-09-954-531-902	Sequence 902, App
C 17	131.8	26.4	360	9	US-10-086-623-1	Sequence 1, Appli
C 18	131.8	26.4	360	9	US-10-260-539-1	Sequence 1, Appli
C 19	74.2	14.8	1472	9	US-10-039-847A-3	Sequence 3, Appli

Sequence 3, Appli
Sequence 3, Appli
Sequence 3, Appli
Sequence 3, Appli
Sequence 13, Appli
Sequence 13, Appli
Sequence 13, Appli
Sequence 13, Appli
Sequence 13, Appli
Sequence 3, Appli
Sequence 3, Appli
Sequence 805, App
Sequence 1238, Ap
Sequence 177, App
Sequence 4, Appli
Sequence 3355, Ap
Sequence 6381, Ap
Sequence 4637, Ap
Sequence 5333, Ap
Sequence 8, Appli
Sequence 14, Appli
Sequence 60, Appli
Sequence 9, Appli
Sequence 27, Appli

US-10-226-559-3
US-10-274-638-3
US-09-808-972-3
US-10-086-623-3
US-10-260-539-3
US-10-076-622-13
US-09-551-621-13
US-09-604-287A-13
US-09-339-338-13
US-10-007-805-13
US-10-224-562-3
US-09-801-861-3
US-09-954-531-805
US-09-954-531-1238
US-10-239-676-177
US-09-754-853A-4
US-09-938-842A-3355
US-10-198-846-6381
US-09-754-853A-4
US-09-960-352-4637
US-09-938-842A-5333
US-10-158-160A-8
US-10-158-160A-14
US-10-074-045-60
US-10-087-464-9
US-09-791-279-27

ALIGNMENTS

RESULT 1
US-10-086-623-5/C
Sequence 5, Application US/10086623
Patent No. US20020164710A1
GENERAL INFORMATION:
APPLICANT: ERIKSSON, Ulf
APPLICANT: AASE, Karin
APPLICANT: LI, Xuri
APPLICANT: PONTEN, Annica
APPLICANT: UUTELA, Marko
APPLICANT: ALITALO, Kari
APPLICANT: OESTMAN, Arne
APPLICANT: HELDIN, Carl-Henrik
TITLE OF INVENTION: PLATELET DERIVED GROWTH FACTOR D, DNA CODING THEREFOR AND USES
FILE REFERENCE: 1064/44833C2
CURRENT APPLICATION NUMBER: US/10/086,623
PRIOR FILING DATE: 2000-03-04
PRIOR FILING DATE: 1998-11-10
PRIOR FILING DATE: 1998-11-10
PRIOR FILING DATE: 1998-12-28
PRIOR FILING DATE: 1999-08-26
PRIOR FILING DATE: 1999-08-26
PRIOR FILING DATE: 1999-10-04
PRIOR FILING DATE: 1999-10-05
PRIOR FILING DATE: 1999-11-10
PRIOR FILING DATE: 2000-10-19
NUMBER OF SEQ ID NOS: 42
SOFTWARE: PatentIn version 3.1
SEQ ID NO 5
LENGTH: 1934
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(966)
OTHER INFORMATION:
US-10-086-623-5

Query Match	100.0%;	Score 500;	DB 9;	Length 1934;..
Best Local Similarity	100.0%;	Pred. No. 3.8e-99;		
Matches 500;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	GTATATGTAAGAAGCCTCATCTTTTGATTTTTAAATATACAAGATGCTTTCTTTAAGAGA	60	
Db	1493	GTATATGTAAGAAGCCTCATCTTTTGATTTTTAAATATACAAGATGCTTTCTTTAAGAGA	1434	
QY	61	GCAAGATTCAAAATTTGTTTGTCTTTCAAAATTTAAAAATAAATTTATCTCCTAAATTTT	120	
Db	1433	GCAAGATTCAAAATTTGTTTGTCTTTCAAAATTTAAAAATAAATTTATCTCCTAAATTTT	1374	
QY	121	CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAGGATTTTAAAGAGTCT	180	
Db	1373	CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAGGATTTTAAAGAGTCT	1314	
QY	181	AACTCAACATATGTAAAGCTCTGGTGACTCTGTTATATATACCAAAAAAACAATTTGAT	240	
Db	1313	AACTCAACATATGTAAAGCTCTGGTGACTCTGTTATATATACCAAAAAAACAATTTGAT	1254	
QY	241	CTATATACATATAGACATGAATATATTTCTGTGTGTGTTTGTCATATATAAAGCTCAAC	300	
Db	1253	CTATATACATATAGACATGAATATATTTCTGTGTGTGTTTGTCATATATAAAGCTCAAC	1194	
QY	301	ACTATTATTAATGCAATCCTATATTTCTTAGGTATAGAAGTTGATGATATACCTTTCTAC	360	
Db	1193	ACTATTATTAATGCAATCCTATATTTCTTAGGTATAGAAGTTGATGATATACCTTTCTAC	1134	
QY	361	TTGCCATGGCATTTAAACAAGCAAGGCTGACACTCAGCAACCACTTGTGTTTCATTGCATTG	420	
Db	1133	TTGCCATGGCATTTAAACAAGCAAGGCTGACACTCAGCAACCACTTGTGTTTCATTGCATTG	1074	
QY	421	CAGGCTAGTAGTAAGTTTGGTTGCTGGTAGGAAAAGGTCCTTATCTCACCTCCCTTAA	480	
Db	1073	CAGGCTAGTAGTAAGTTTGGTTGCTGGTAGGAAAAGGTCCTTATCTCACCTCCCTTAA	1014	
QY	481	ACTAAAGGTTCTTTTCAGGCT	500	
Db	1013	ACTAAAGGTTCTTTTCAGGCT	994	

~~RESULT 2~~

RESOLUT 2
US-10-260-539-5/c
; Sequence 5, Application US/10260539
; Publication No. US20030073637A1
; GENERAL INFORMATION:
; APPLICANT: ERIKSSON, Ulf
; APPLICANT: AASE, Karin
; APPLICANT: LI, Xuri
; APPLICANT: PONTEN, Annica
; APPLICANT: UUTELA, Marko
; APPLICANT: ALITALO, Kari
; APPLICANT: OESTMAN, Arne
; APPLICANT: HELDIN, Carl-Henrik
; TITLE OF INVENTION: PLATELET DERIVED GROWTH FACTOR D, DNA CODING THEREFOR AND USES THEREOF
; FILE REFERENCE: 1064/44833C2
; CURRENT APPLICATION NUMBER: US/10/260,539
; CURRENT FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: US/10/086,623
; PRIOR FILING DATE: 2000-03-04
; PRIOR APPLICATION NUMBER: US 60/107,852
; PRIOR FILING DATE: 1998-11-10
; PRIOR APPLICATION NUMBER: US 60/113,997
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: US 60/150,604
; PRIOR FILING DATE: 1999-08-26
; PRIOR APPLICATION NUMBER: US 60/157,108
; PRIOR FILING DATE: 1999-10-04
; PRIOR APPLICATION NUMBER: US 60/157,756
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: US 09/438,046
; PRIOR FILING DATE: 1999-11-10

```

; PRIOR APPLICATION NUMBER: US 09/691,200
;
; PRIOR FILING DATE: 2000-10-19
;
; NUMBER OF SEQ ID NOS: 42
;
; SOFTWARE: PatentIn version 3.1
;
; SEQ ID NO 5
;
; LENGTH: 1934
;
; TYPE: DNA
;
; ORGANISM: Homo sapiens
;
; FEATURE:
;
; NAME/KEY: CDS
;
; LOCATION: (1)..(966)
;
; OTHER INFORMATION:
;
; US-10-260-539-5

```

Query Match	100.0%;	Score 500;	DB 9;	Length 1934;
Best Local Similarity	100.0%;	Pred. No. 3.8e-99;		
Matches 500; Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
Qy	1	GTATATCTAGAAGCCTCATCTTTTGATTTTAAATATACAAGATGCTTCTTTAAGAGA	60	
Db	1493	GTATATGTAAGAAGCCTCACTCTTTTGATTTTAAATATACAAGATGCTTCTTTAAGAGA	1434	
Qy	61	GCAAGATTCAAAATTTGTTTGCTTTTCAAAATTTAAAAATAAAATTTATCTCCTFAAATTTT	120	
Db	1433	GCAAGATTCAAAATTTGTTTGCTTTTCAAAATTTAAAAATAAAATTTATCTCCTFAAATTTT	1374	
Qy	121	CTAAGACATGTTTCATATATATTTGACATCCCTTATTTTGGCAAGAGATTTTAAAGAGTCT	180	
Db	1373	CTAAGACATGTTTCATATATATTTGACATCCCTTATTTTGGCAAGAGATTTTAAAGAGTCT	1314	
Qy	181	AACCTCAACATATGTAAAGCTCTGGTCTGCTGCTATATATATACCAAAAAAACAATTTGAT	240	
Db	1313	AACCTCAACATATGTAAAGCTCTGGTCTGCTGCTATATATATACCAAAAAAACAATTTGAT	1254	
Qy	241	CTATATACACATAGACATGAATATATTTCTGCTGCTGTTGTGCATATATAAACCCTCAAAC	300	
Db	1253	CTATATACACATAGACATGAATATATTTCTGCTGCTGTTGTGCATATATAAACCCTCAAAC	1194	
Qy	301	ACTATTATTTAAATGCAATCCCTATATTTCTTAGGTATAGAGTTGATGATATACCTTTCTAC	360	
Db	1193	ACTATTATTTAAATGCAATCCCTATATTTCTTAGGTATAGAGTTGATGATATACCTTTCTAC	1134	
Qy	361	TTGCCATGGCATTTAAACAAGCAAGGCTGAGACTCAGCAACCACTGTGTGTCATTGCAATG	420	
Db	1133	TTGCCATGGCATTTAAACAAGCAAGGCTGAGACTCAGCAACCACTGTGTGTCATTGCAATG	1074	
Qy	421	CAGGCTAGTAGTAAGTTTGGTTTGGCTGGTAGGAAAAGGTCCTTTATCTCACCCCTCCTTAA	480	
Db	1073	CAGGCTAGTAGTAAGTTTGGTTTGGCTGGTAGGAAAAGGTCCTTTATCTCACCCCTCCTTAA	1014	
Qy	481	ACTAAAGGTTCTTTTCAGGCT 500		
Db	1013	ACTAAAGGTTCTTTTCAGGCT 994		

RESULT 3

US-10-086-623-7/c
 Sequence 7, Application US/10086623
 Patent No. US20020164710A1
 GENERAL INFORMATION:
 APPLICANT: ERIKSSON, Ulf
 APPLICANT: NASE, Karin
 APPLICANT: Li, Xuri
 APPLICANT: PONTEN, Annica
 APPLICANT: UUTELA, Marko
 APPLICANT: ALITALO, Kari
 APPLICANT: OESTMAN, Arne
 APPLICANT: HELDIN, Carl-Henrik
 TITLE OF INVENTION: PLATELET DERIVED GROWTH FACTOR D, DNA CODING THEREFOR AND USES
 FILE REFERENCE: 1064/4833C2
 CURRENT APPLICATION NUMBER: US/10/086,623
 CURRENT FILING DATE: 2000-03-04
 PRIOR APPLICATION NUMBER: US 60/107,852

~~RESULT 4~~
US-10-260-539-7/c
; Sequence 7, Application US/10260539
; Publication No. US20030073637A1
; GENERAL INFORMATION:
; APPLICANT: ERIKSSON, Ulf
; APPLICANT: AASE, Karin

Qy	1	GTATATGTAGAAGCCTCATCTTTTGATTTTAAATATACAAGATGCTTCTTTAAGAGA	60
Db	1812	GTATATGTAGAAGCCTCATCTTTTGATTTTAAATATACAAGATGCTTCTTTAAGAGA	1753
Qy	61	GCAAGATTCAAAATTTGTTTGTCTTTCACAAATTTAAAAATAAATTTATCTCCCTAAATTTT	120
Db	1752	GCAAGATTCAAAATTTGTTTGTCTTTCACAAATTTAAAAATAAATTTATCTCCCTAAATTTT	1693
Qy	121	CTAAGACATGTTTCATATATTTGACATCCCTTATTTGGCAAGGATTTTAAAGAGTCT	180
Db	1692	CTAAGACATGTTTCATATATTTGACATCCCTTATTTGGCAAGGATTTTAAAGAGTCT	1633
Qy	181	AACTCAACATATGTAGCTCTGGTGACTGCTTATATATACCAAAAAAACAATTTGAT	240
Db	1632	AACTCAACATATGTAGCTCTGGTGACTGCTTATATATACCAAAAAAACAATTTGAT	1573
Qy	241	CTATATACATAGACATGAATATATTTCTGTGTGTGTTGTGCATATATAACCTCAAC	300
Db	1572	CTATATACATAGACATGAATATATTTCTGTGTGTGTTGTGCATATATAACCTCAAC	1513
Qy	301	ACTATTATTAATGCAATCCTATATTCCTTAGGTATAGAAGTTGATGATATACCTTCTAC	360
Db	1512	ACTATTATTAATGCAATCCTATATTCCTTAGGTATAGAAGTTGATGATATACCTTCTAC	1453
Qy	361	TTGCCATGGCATTTAAACAAACCAAGGCTGAGACTCAGCAACCACTTGTGTTTCATTGCATTG	420
Db	1452	TTGCCATGGCATTTAAACAAACCAAGGCTGAGACTCAGCAACCACTTGTGTTTCATTGCATTG	1393
Qy	421	CAGGCTAGTAGTAAGTTTGGTTGCTGGTAGGAAAAAGGTCCTTATCTACCCCTCCTTAA	480
Db	1392	CAGGCTAGTAGTAAGTTTGGTTGCTGGTAGGAAAAAGGTCCTTATCTACCCCTCCTTAA	1333

✓
RESULT 7

US-10-083-853-1
; Sequence 1, Application US/10083853
; Patent No. US20020164709A1
; GENERAL INFORMATION:
; APPLICANT: Affymetrix, Inc
; APPLICANT: Shigeta, Ron T
; APPLICANT: Siani-Rose, Michael A
; TITLE OF INVENTION: Nucleic Acid Encoding Growth Factor Protein
; FILE REFERENCE: 3385.1
; CURRENT APPLICATION NUMBER: US/10/083.853
; PRIOR FILING DATE: 2002-02-26
; PRIOR APPLICATION NUMBER: USN 60/272.663
; PRIOR FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 29921
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-10-083-853-1

Query Match 100.0%; Score 500; DB 9; Length 29921;
Best Local Similarity 100.0%; Pred. No. 1.2e-98;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTATATGTAAGAAGCCCTCATCTTTTGATTTTAAATATACAAGATGCTTTCTTTAAAGAGA 60
|||||
DB 1 GTATATGTAAGAAGCCCTCATCTTTTGATTTTAAATATACAAGATGCTTTCTTTAAAGAGA 60
|||||

QY 61 GCAAGATTCAAAATGTTTGTGTTTCAAAATTTAAATTTAAATTTATCTCTAAATTTT 120
|||||
DB 61 GCAAGATTCAAAATGTTTGTGTTTCAAAATTTAAATTTAAATTTATCTCTAAATTTT 120
|||||

QY 121 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAGAGATTTTAAAGAGTCT 180
|||||
DB 121 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAGAGATTTTAAAGAGTCT 180
|||||

QY 181 AACTCAACATATGTAAGCTCTGTGTACCTGGTTATATATACCAAAAAACATTTGAT 240
|||||
DB 181 AACTCAACATATGTAAGCTCTGTGTACCTGGTTATATATACCAAAAAACATTTGAT 240
|||||

QY 241 CTATATACATAGACATGAATATTTCTGTGTGTTTCTGTGATATATAACCTTCAAC 300
|||||
DB 241 CTATATACATAGACATGAATATTTCTGTGTGTTTCTGTGATATATAACCTTCAAC 300
|||||

QY 301 ACTATTATTAATGCAATCCCTATATTTCTAGGTATAGAAGTTGATGATATACCTTTCTAC 360
|||||
DB 301 ACTATTATTAATGCAATCCCTATATTTCTAGGTATAGAAGTTGATGATATACCTTTCTAC 360
|||||

QY 361 TTGCATGGCATTAACAAGCAAGCTGAGACTCAGCAACACCTTGTTGTTGATGATGATG 420
|||||
DB 361 TTGCATGGCATTAACAAGCAAGCTGAGACTCAGCAACACCTTGTTGTTGATGATGATG 420
|||||

QY 421 CAGGCTAGTAGTAAGTTGGTTGTTGGTGGAGAAAGGCTCTCTTATCTCACCCCTCTTAA 480
|||||
DB 421 CAGGCTAGTAGTAAGTTGGTTGTTGGTGGAGAAAGGCTCTCTTATCTCACCCCTCTTAA 480
|||||

QY 481 ACTAAAGGTTCTTTCAGGCT 500
|||||
DB 481 ACTAAAGGTTCTTTCAGGCT 500
|||||

RESULT 8

US-09-915-582-35/c
; Sequence 35, Application US/09915582
; Patent No. US20020120103A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 17 Human Secreted Proteins
; FILE REFERENCE: PS723P1

; CURRENT APPLICATION NUMBER: US/09/915.582
; CURRENT FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: PCT/US01/01431
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/231,968
; PRIOR FILING DATE: 2000-09-12
; NUMBER OF SEQ ID NOS: 97
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 35
; LENGTH: 4001
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-915-582-35

Query Match 99.9%; Score 499.6; DB 10; Length 4001;
Best Local Similarity 99.8%; Pred. No. 6.2e-99;
Matches 499; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTATATGTAAGAAGCCCTCATCTTTTGATTTTAAATATACAAGATGCTTTCTTTAAAGAGA 60
|||||
DB 1999 GTATATGTAAGAAGCCCTCATCTTTTGATTTTAAATATACAAGATGCTTTCTTTAAAGAGA 1940
|||||

QY 61 GCAAGATTCAAAATGTTTGTGTTTCAAAATTTAAATTTAAATTTATCTCTAAATTTT 120
|||||
DB 1939 GCAAGATTCAAAATGTTTGTGTTTCAAAATTTAAATTTAAATTTATCTCTAAATTTT 1880
|||||

QY 121 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAGAGATTTTAAAGAGTCT 180
|||||
DB 1879 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAGAGATTTTAAAGAGTCT 1820
|||||

QY 181 AACTCAACATATGTAAGCTCTGTGTACCTGGTTATATATACCAAAAAACATTTGAT 240
|||||
DB 1819 AACTCAACATATGTAAGCTCTGTGTACCTGGTTATATATACCAAAAAACATTTGAT 1760
|||||

QY 241 CTATATACATAGACATGAATATTTCTGTGTGTTTCTGTGATATATAACCTTCAAC 300
|||||
DB 1759 CTATATACATAGACATGAATATTTCTGTGTGTTTCTGTGATATATAACCTTCAAC 1700
|||||

QY 301 ACTATTATTAATGCAATCCCTATATTTCTAGGTATAGAAGTTGATGATATACCTTTCTAC 360
|||||
DB 1699 ACTATTATTAATGCAATCCCTATATTTCTAGGTATAGAAGTTGATGATATACCTTTCTAC 1640
|||||

QY 361 TTGCATGGCATTAACAAGCAAGCTGAGACTCAGCAACACCTTGTTGTTGATGATGATG 420
|||||
DB 1639 TTGCATGGCATTAACAAGCAAGCTGAGACTCAGCAACACCTTGTTGTTGATGATGATG 1580
|||||

QY 421 CAGGCTAGTAGTAAGTTGGTTGTTGGTGGAGAAAGGCTCTCTTATCTCACCCCTCTTAA 480
|||||
DB 1579 CAGGCTAGTAGTAAGTTGGTTGTTGGTGGAGAAAGGCTCTCTTATCTCACCCCTCTTAA 1520
|||||

QY 481 ACTAAAGGTTCTTTCAGGCT 500
|||||
DB 1519 ACTAAAGGTTCTTTCAGGCT 1500
|||||

RESULT 9
US-09-915-582-17/c
; Sequence 17, Application US/09915582
; Patent No. US20020120103A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 17 Human Secreted Proteins
; FILE REFERENCE: PS723P1
; CURRENT APPLICATION NUMBER: US/09/915.582
; CURRENT FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: PCT/US01/01431
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31

Qy

61 GCACGATTCAAATTTGTTTGTGTTTCAAAAATTTAAAAATTAATTTATCTCCTAAATTTT 120

Db 1746 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAGGATTTTAAGAGTCT 1687
QY 181 AACTCAAAACATATGTAAGCTCTGTTGACCTGGTTATATATACC----AAAAAAACATT 236
|||||
Db 1686 AACTCAAAACATATGTAACCTCTGTTGACCTGGTTATATATACCAAAAAAAACATT 1627
|||||
QY 237 TGATCTATATACACATAGACATGAATATATTTCTGTGTGTGTTTGTGCATATATAACCTC 296
|||||
Db 1626 TGATCTATATACACATAGACATGAATATATTTCTGTGTGTGTTTGTGCATATATAACCTC 1567
|||||
QY 297 AAACACTATTTAAATGCAATCCTATATTTCTTAGGTATAGAAAGTTGATGATATACCTTT 356
|||||
Db 1566 AAACACTATTTAAATGCAATCCTATATTTCTTAGGTATAGAAAGTTGATGATATACCTTT 1507
|||||
QY 357 CTACTTGCCATGGCATTAAACAAGCAAGGCTGAGACTCAGCAACCACTTGTGTTCAATTGC 416
|||||
Db 1506 CTACTTGCCATGGCATTAAACAAGCAAGGCTGAGACTCAGCAACCACTTGTGTTCAATTGC 1447
|||||
QY 417 ATTGCAGGCTAGTAGTAAGTTTGGTTGCTGCTAGGAAAAGGTTCTCTTATCTCACCCCTCC 476
|||||
Db 1446 ATTGCAGGCTAGTAGTAAGTTTGGTTGCTGCTAGGAAAAGGTTCTCTTATCTCACCCCTCC 1387
|||||
QY 477 TTAAACTAAAGGTTCTTTCAGGCT 500
|||||
Db 1386 TTAAACTAAAGGTTCTTTCAGGCT 1363
|||||

Search completed: July 4, 2003, 19:36:11
Job time : 113.851 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 15:08:57 ; Search time 1508.67 Seconds
(without alignments)
9664.474 Million cell updates/sec

Table: us-10-083-853b-2_copy_29421_29921
Perfect score: 501
Sequence: 1 ctttggttttagacatg.....caattctgtgaagaagtaa 501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database 8 GenBank

1: gb-ba.*

2: gb-htg.*

3: gb-in.*

4: gb-om.*

5: gb-lov.*

6: gb-pat.*

7: gb-ph.*

8: gb-pl.*

9: gb-pr.*

10: gb-ro.*

11: gb-sts.*

12: gb-sy.*

13: gb-un.*

14: gb-vi.*

15: em-ba.*

16: em-fun.*

17: em-hum.*

18: em-in.*

19: em-mu.*

20: em-or.*

21: em-ov.*

22: em-pat.*

23: em-ph.*

24: em-pl.*

25: em-ro.*

26: em-sts.*

27: em-un.*

28: em-vi.*

29: em-htg-hum.*

30: em-htg-inv.*

31: em-htg-other.*

32: em-htg-mus.*

33: em-htg-pln.*

34: em-htg-rod.*

35: em-htg-mam.*

36: em-htg-vrt.*

37: em-sy.*

38: em-htg-hum.*

39: em-htg-mus.*

40: em-htg-other.*

41: em-htg-other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	501	100.0	180648	9	AP003043	AP003043 Homo sapi
2	501	100.0	182403	2	AC024052	AC024052 Homo sapi
3	480.2	95.8	46442	9	AL591394	AL591394 Human DNA
4	478.6	95.5	47364	9	AC008809	AC008809 Homo sapi
5	478.6	95.5	206163	9	AC099519	AC099519 Homo sapi
6	478.6	95.5	206865	2	AC027027	AC027027 Homo sapi
7	477	95.2	94727	9	AL591034	AL591034 Human DNA
8	477	95.2	149041	9	AC044782	AC044782 Homo sapi
9	477	95.2	154076	2	AC023273	AC023273 Homo sapi
10	477	95.2	162841	9	AC099504	AC099504 Homo sapi
11	477	95.2	169161	2	AC019053	AC019053 Homo sapi
12	477	95.2	171490	2	AC068209	AC068209 Homo sapi
13	477	95.2	176630	2	AC010684	AC010684 Homo sapi
14	477	95.2	179789	9	AC069843	AC069843 Homo sapi
15	477	95.2	184000	9	AC087774	AC087774 Homo sapi
16	475.4	94.9	3994	9	AC111191	AC111191 Homo sapi
17	475.4	94.9	41260	9	AC114309	AC114309 Homo sapi
18	475.4	94.9	116800	9	AL391379	AL391379 Human DNA
19	475.4	94.9	142391	2	AC092507	AC092507 Homo sapi
20	475.4	94.9	147423	9	AC099487	AC099487 Homo sapi
21	475.4	94.9	150572	9	CNS01RIC	AL163151 Human chr
22	475.4	94.9	151018	9	AC096756	AC096756 Homo sapi
23	475.4	94.9	152741	2	AC021606	AC021606 Homo sapi
24	475.4	94.9	156485	9	AC007422	AC007422 Homo sapi
25	475.4	94.9	163673	2	AC027598	AC027598 Homo sapi
26	475.4	94.9	165295	2	AC068486	AC068486 Homo sapi
27	475.4	94.9	165714	9	AC016919	AC016919 Homo sapi
28	475.4	94.9	168063	9	AC080051	AC080051 Homo sapi
29	475.4	94.9	172543	9	AC106714	AC106714 Homo sapi
30	475.4	94.9	172805	9	AC023134	AC023134 Homo sapi
31	475.4	94.9	173288	9	AC104167	AC104167 Homo sapi
32	475.4	94.9	176624	2	AC104394	AC104394 Homo sapi
33	475.4	94.9	176903	9	AC008501	AC008501 Homo sapi
34	475.4	94.9	178402	2	AC090508	AC090508 Homo sapi
35	475.4	94.9	178780	9	AF216667	AF216667 Homo sapi
36	475.4	94.9	180339	9	AC099049	AC099049 Homo sapi
37	475.4	94.9	181236	2	AC022757	AC022757 Homo sapi
38	475.4	94.9	183437	9	AC068933	AC068933 Homo sapi
39	475.4	94.9	189053	2	AC016015	AC016015 Homo sapi
40	475.4	94.9	190195	9	AC016954	AC016954 Homo sapi
41	475.4	94.9	192586	9	AC018437	AC018437 Homo sapi
42	475.4	94.9	194867	2	AC023152	AC023152 Homo sapi
43	475.4	94.9	195931	2	AC128693	AC128693 Homo sapi
44	475.4	94.9	196660	9	AC007000	AC007000 Homo sapi
45	475.4	94.9	200466	2	AL845433	AL845433 Homo sapi

ALIGNMENTS

RESULT 1
AP003043
LOCUS
DEFINITION Homo sapiens genomic DNA, chromosome 11q clone:RP11-617B3, complete sequence.
ACCESSION AP003043
VERSION AP003043.2 GI:15637094
KEYWORDS HTG.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE
1 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
linear PR1 18-SEP-2001
clone:RP11-617B3, complete

TITLE Homo sapiens genomic DNA
JOURNAL Published Only in Database (2000)
REFERENCE 2 (bases 1 to 180648)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (12-DEC-2000) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT On Sep 17, 2001 this sequence version replaced gi:11862945.
FEATURES
source
1..180648
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-617B3"
BASE COUNT 56234 a 35375 c 33753 g 55286 t
ORIGIN

Query Match 100.0%; Score 501; DB 9; Length 180648;
Best Local Similarity 100.0%; Pred. No. 5e-112;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CTTTGGTGTGTAGACATCAAGTCCCTGCGCCATGCTATGCTGTAATGGTATTCGCTG 60
Db CTTTGGTGTGTAGACATCAAGTCCCTGCGCCATGCTATGCTGTAATGGTATTCGCTG 69350
Qy 61 GGTTCCTCTAGGATTTTATGGTTTAAAGTCTAAATTAAGTCTTTAATCCATCTTG 120
Db GGTTCCTCTAGGATTTTATGGTTTAAAGTCTAAATTAAGTCTTTAATCCATCTTG 69410
Qy 121 AATTAATTTTATAGGTGTAGAGGATCCAGTTTCAGCTTCTACATAGGCTA 180
Db AATTAATTTTATAGGTGTAGAGGATCCAGTTTCAGCTTCTACATAGGCTA 69470
Qy 181 GCCAGTTTCTCAGCACCATTTAATAAGGAATCCTTCCCATGCTGTTCTTCT 240
Db GCCAGTTTCTCAGCACCATTTAATAAGGAATCCTTCCCATGCTGTTCTTCT 69530
Qy 241 CAGTTTGTCAAGATCAGATAGTTGTAGATATCGCGCATATTCTCGAGGCTCTGTT 300
Db CAGTTTGTCAAGATCAGATAGTTGTAGATATCGCGCATATTCTCGAGGCTCTGTT 69590
Qy 301 TGTTCATTCGTTGATATCTCTTTTGGTACCAGTACCAGTGTGTTGGTTACTGTAG 360
Db TGTTCATTCGTTGATATCTCTTTTGGTACCAGTACCAGTGTGTTGGTTACTGTAG 69650
Qy 361 CCTGTAGTGTAGTTGAAGTCAGGTAGCATGATGCTCCAGCTTTGTTCTTTGGCTTA 420
Db CCTGTAGTGTAGTTGAAGTCAGGTAGCATGATGCTCCAGCTTTGTTCTTTGGCTTA 69710
Qy 421 GGATTTGCTGGCATGTGGGCTCTTTTGGTTCCACATGAATTAAGTATGTTTTT 480
Db GGATTTGCTGGCATGTGGGCTCTTTTGGTTCCACATGAATTAAGTATGTTTTT 69770
Qy 481 CCAATTCGTGAAGAAGTAA 501
Db CCAATTCGTGAAGAAGTAA 69791

RESULT 2
AC024052 182403 bp DNA linear HTG 17-AUG-2000
LOCUS Homo sapiens chromosome 11 clone RP11-617B3, WORKING DRAFT
DEFINITION SEQUENCE, 7 unordered pieces.
AC024052
AC024052.3 GI:9838295
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 182403)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 182403)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (20-FEB-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT On Aug 17, 2000 this sequence version replaced gi:7109658.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0617B03
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: Plasmid; 0%
Chemistry: Dye-primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 178574 bases at least Q40
Consensus quality: 179480 bases at least Q30
Consensus quality: 179940 bases at least Q20
Insert size: 198000; agarose-fp
Insert size: 183143; sum-of-contigs
Quality coverage: 5.39 in Q20 bases; agarose-fp
Quality coverage: 5.87 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 7 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
* 1 1666: contig of 1666 bp in length
* 1667 1766: gap of unknown length
* 1767 9133: contig of 7367 bp in length
* 9134 9233: gap of unknown length
* 9234 20663: contig of 11430 bp in length
* 20664 20763: gap of unknown length
* 20764 35927: contig of 15164 bp in length
* 35928 36028: gap of unknown length
* 36029 61610: contig of 25583 bp in length
* 61611 61710: gap of unknown length
* 61711 96167: contig of 34457 bp in length
* 96168 96268: gap of unknown length
* 96269 182403: contig of 86136 bp in length.
FEATURES
source
1..182403
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/clone="RP11-617B3"
/notes="assembly_name:Contig2"
1..1666
/notes="assembly_name:Contig3"
1767..9133
/notes="assembly_name:Contig3"
clone_end:SF6
vector_side:left
9234..20663
/notes="assembly_name:Contig4"
20764..35927
/notes="assembly_name:Contig5"
36028..61610
/notes="assembly_name:Contig6"

corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at

Database can be found at: http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/ChrX>
RP11-1145B22 is from the library RP11-11.4 constructed by the group
of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>

VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-1145B22 It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone RP1-54B20 is at 44443 in this sequence.
The true right end of clone RP3-393P12 is at 2000 in this sequence.

FEATURES
source

source

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="p11.21-11.23"
/clone="RP11-1145B22"
/clone_lib="pPCT-11.4"
```

Query Match	95.98;	Score 480.2;	DB 9;	Length 46442;
Best Local Similarity	97.4%;	Pred. No. 6.6e-107;		
Matches 488;	Conservative 0;	Mismatches 13;	Indels 0;	Gaps 0;
Qy 1	CTTTTGGTGTTTAGACATGAAGTCCTTGCCCATGCTATGCTCCTGAATGGTATTGCCTG	60		
Db 41112	CTTTTGGTGTTTAGACATGAAGTCCTTGCCCGCTGCTATGCTCCTGAATGGTATTGCCTA	4105		
Qy 61	GGTTTTCTCTACGGATTTTATGGTTTAGGTTAGGTTCTAACATTTAAGTCTTTAATCCATCTTG	120		
Db 41052	AGTTTTCTCTACGGATTTTATGGTTTAGGTTAGGTTCTAACATTTAAGTCTTTAATCCATCTTG	4099		
Qy 121	AATTAATTTTGTATAAGGTGTAAAGGAAGGATCCAGTTTCAGCTTTCTACATAGGGCTA	180		
Db 40992	AATTAATTTTGTATAAGGTGTAAAGGAAGGATCCAGTTTCAGCTTTCTACATAGGGCTA	4093		
Qy 181	GCAGTTTTCTCAGCACCATTATTAATAAGGGAATCCCTTCCCATGCTGTTTTCT	240		
Db 40932	GCAGTTTTCCAGCACCATTATTAATAAGGGAATCCCTTCCCATGCTGTTTTCT	4087		
Qy 241	CAGGTTTGTCAAAGATCAGATAGTTGTAGATATGCGGCATATTCTCAGGGGCTCTGTTCT	300		
Db 40872	CAGGTTTGTCAAAGATCAGATAGTTGTAGATATGCGGCATATTCTCAGGGGCTCTGTTCT	4081		
Qy 301	TGTTCCATTTGGTTGATATCTGTTTTGGTACCAGTACCATGTTGTTTTGGTTACGTAG	360		
Db 40812	TGTTCCATTTGGTCTATATCTGTTTTGGTACCAGTACCATGCTGTTTTGGTTACTGTAG	4075		
Qy 361	CTTTGTAGTCTAGTTTGAAGTTCAGGTAGCATGATGCTCCAGCTTTGTTCTTTTGGCTTA	420		
Db 40752	CTTTGTAGTATAGTTTGAAGTTCAGGTAGCATGATGCTCCAGCTTTGTTCTTTTGGCTTA	4069		
Qy 421	GGATTGACTTGGCGATGTGGGCTCTTTTTTTGGTTCCACATGAACACTTTAAAGTAGTTTTT	480		
Db 40692	GGATTGACTTGGCGATGTGGGCTCTTTTTTTGGTTCCATATGAACACTTTAAAGTAGTTTTT	4063		

[illegible]

QY	181	GCCAGTTTTCTCAGCACCAATTTATTAAATAGGGAATCCTTCCCATTGCCTGTGTTTTCT	240
Db	86373	GCACGTTTTCCAGCAGCACCATTTATTAAATAGGGAATCCTTCCCATTGCCTGTGTTTTCT	86314
QY	241	CAGGTTTTGTCAAAGATCAGATAGTTGTAGATATCGGCATTAATTTCTCAGGGCTCTGTTTC	300
Db	86313	CAGGTTTTGTCAAAGATCAGATAGTTGTAGATATCGGCATTAATTTCTCAGGGCTCTGTTTC	86254
QY	301	TGTTCCATTTGGTTTGATATCTCTGTTTTGGTAGCAGTACCATGTTGTTTGGTTACTGTAG	360
Db	86253	CGTTCCATTTGGTCTATATCTCTGTTTTGGTAGCAGTACCATGTTGTTTGGTTACTGTAG	86194
QY	361	CCTTGTAGTGTAGTTTGAAGTCAGGTACGATGCCCTCCAGCTTGTCTTTTGGGCTTA	420
Db	86193	CCTTGTAGTATAGTTTGAAGTCAGGTACGATGCCCTCCAGCTTGTCTTTTGGGCTTA	86134
QY	421	GGATTGACTTGGCGATGTGGGCTCTTTTTTGGTCCACATGAACCTTTAAAGTAGTTTTTT	480
Db	86133	GGATTGACTTGGTGTAGTGTGGGCTCTTTTTTGGTCCATATGNACTTTAAAGTAGTTTTTT	86074
QY	481	CCAATTCTGTGAAGAAGTAA	501
Db	86073	CCAATTCTGTGAAGAAGTCA	86053

RESULT 6	AC027027	206865 bp	DNA	linear	HTG 25-MAY-2000
LOCUS	AC027027/c				
DEFINITION	Homo sapiens chromosome 5 clone RP11-695010 map 5, WORKING DRAFT SEQUENCE, 19 unordered pieces.				
ACCESSION	AC027027				
VERSION	AC027027.3	GI:8081300			
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.				
SOURCE	Homo sapiens.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 206865)				
TITLE	Birren,B., Linton,L., Nusbaum,C. and Lander,E.				
JOURNAL	Homo sapiens chromosome 5, clone RP11-695010				
REFERENCES	Unpublished				
AUTHORS	2 (bases 1 to 206865)				
	Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bada,F., Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Campiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrum,J., Meneses,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severi,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.				
TITLE	Direct Submission				
JOURNAL	Submitted (25-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA				
COMMENT	On May 25, 2000 this sequence version replaced gi:7677916. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html				
	----- Genome Center				
	Center: Whitehead Institute/ MIT Center for Genome Research				

misc_feature 1115..3783
/note="assembly_fragment"
misc_feature 3884..6474
/note="assembly_fragment"
clone_end:T7
vector_side:right
misc_feature 6575..11183
/note="assembly_fragment"
misc_feature 11284..16799
/note="assembly_fragment"
misc_feature 16900..21561
/note="assembly_fragment"
misc_feature 21662..27796
/note="assembly_fragment"
misc_feature 27897..33998
/note="assembly_fragment"
misc_feature 34099..41507
/note="assembly_fragment"
misc_feature 41608..47859
/note="assembly_fragment"
clone_end:SP6
vector_side:left
misc_feature 47960..56205
/note="assembly_fragment"
misc_feature 56306..68561
/note="assembly_fragment"
misc_feature 68662..81542
/note="assembly_fragment"
misc_feature 81643..96660
/note="assembly_fragment"
misc_feature 96761..114606
/note="assembly_fragment"
misc_feature 114707..136359
/note="assembly_fragment"
misc_feature 136460..156648
/note="assembly_fragment"
misc_feature 156749..178778
/note="assembly_fragment"
misc_feature 178879..206865
/note="assembly_fragment"
BASE COUNT 64815 a 37677 c 37421 g 65147 t 1805 others
ORIGIN

Query Match 95.5%; Score 478.6; DB 2; Length 206865;
Best Local Similarity 97.2%; Pred. No. 1.4e-106;
Matches 487; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 CTTTGGTGTATAGACATGAGTCCTGCGCATGCTATGCTGATGCTGCTG 60
DB 71931 CTTTGGTGTATAGACATGAGTCCTGCGCATGCTATGCTGATGCTGCTG 71872
QY 61 GGTTCCTCTAGGCAATTTATGTTTATGCTCTAACATTTAGTCTTTATCCATCTTG 120
DB 71871 GGTTCCTCTAGGCAATTTATGTTTATGCTCTAACATTTAGTCTTTATCCATCTTG 71812
QY 121 AATTAATTTTGTATAGGTGTATAGGAGGATCCAGTTTCAGCTTTCTACATAGGCTA 180
DB 71811 AATTAATTTTGTATAGGTGTATAGGAGGATCCAGTTTCAGCTTTCTACATAGGCTA 71752
QY 181 GCCAGTTTCTCAGCACCAATTTATTAATAGGGAATCTTCCCATGCTGTTTCT 240
DB 71751 GCCAGTTTCTCAGCACCAATTTATTAATAGGGAATCTTCCCATGCTGTTTCT 71692
QY 241 CAGGTTCCTCAAGATCAGATAGTGTAGATATGCGGCATTTATCTCGAGGCTCTGTC 300
DB 71691 CAGGTTCCTCAAGATCAGATAGTGTAGATATGCGGCATTTATCTCGAGGCTCTGTC 71632
QY 301 TGTTCATTTGGTGAATCTGTTTGGTACAGTACCAATGTTGTTTGGTACTGTAG 360
DB 71631 CGTTCATTTGGTGAATCTGTTTGGTACCAATGTTGTTTGGTACTGTAG 71572
QY 361 CCTGTGTAGTCTGTTTGAATCAGTATGATGCTCCTCAGCTTCTTCTTTGGCTTA 420

Db 71571 CTTTGTAGTATAGTTTGAAGTCAGGTAGCATGATGCTCAGCTTCTTTGGCTTA 71512
QY 421 GGATTGACTTGGCATGTGGCTCTTTTGGTTCACATGAACCTTTAAAGTACTTTTT 480
Db 71511 GGATTGACTTGGCATGTGGCTCTTTTGGTTCACATGAACCTTTAAAGTACTTTTT 71452
QY 481 CCAATTCCTGTGAAGAAAGTAA 501
Db 71451 CCAATTCCTGTGAAGAAAGTCA 71431
RESULT 7
AL591034 94727 bp DNA linear PRI 13-JUN-2001
LOCUS Human DNA sequence from clone RP11-96J19 on chromosome 6, complete
DEFINITION
ACCESSION AL591034
VERSION AL591034.5 GI:14456374
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 94727)
AUTHORS Williams,S.
TITLE Direct Submission
JOURNAL Submitted (13-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
COMMENT On Jun 15, 2001 this sequence version replaced gi:14270021.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; Sw:
SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information on the WormPEP
database can be found at
http://www.sanger.ac.uk/projects/C.elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
RP11-96J19 is from the library RPCI-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-96J19. It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true left end of clone RP1-27K12 is at 92728 in this sequence.
The true right end of clone RP3-483K16 is at 2000 in this sequence.
FEATURES
source
1..94727
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP11-96J19"
/clone_lib="RPCI-11.1"
182..322
/note="L2 repeat: matches 2555..2707 of consensus"
443..588
/note="FLAM_C repeat: matches 1..143 of consensus"
613..732
repeat_region
repeat_region
repeat_region

```
repeat_region /note="Charlie5 repeat: matches 4. .116 of consensus"
repeat_region /note="Charlie5 repeat: matches 2241. .2391 of consensus"
repeat_region /note="L1MB4 repeat: matches 5628. .5767 of consensus"
repeat_region /note="THEIC repeat: matches 1. .371 of consensus"
repeat_region /note="THEIC-internal repeat: matches 500. .1580 of consensus"
repeat_region /note="AluSg repeat: matches 1. .293 of consensus"
repeat_region /note="THEIC-internal repeat: matches 3. .500 of consensus"
repeat_region /note="AluY repeat: matches 1. .311 of consensus"
repeat_region /note="THEIC repeat: matches 1. .371 of consensus"
repeat_region /note="L1MB4 repeat: matches 5764. .6182 of consensus"
repeat_region /note="THE1-INTERNAL repeat: matches 752. .877 of consensus"
misc_feature /note="match: GSS: Em:A2576617"
repeat_region /note="85 copies 2 mer tt 67% conserved"
repeat_region /note="3 copies 49 mer 78% conserved"
repeat_region /note="AluSx repeat: matches 1. .309 of consensus"
repeat_region /note="AluSg repeat: matches 1. .307 of consensus"
repeat_region /note="FLAM_A repeat: matches 1. .133 of consensus"
repeat_region /note="AluSg repeat: matches 1. .308 of consensus"
repeat_region /note="MIR repeat: matches 71. .209 of consensus"
repeat_region /note="L2 repeat: matches 2640. .2749 of consensus"
misc_feature /note="match: GSS: Em:A0106154"
repeat_region /note="AluSg/x repeat: matches 128. .290 of consensus"
misc_feature complement(13683. .14100)
misc_feature /note="match: GSS: Em:AQ265669"
repeat_region /note="match: GSS: Em:A0683800"
misc_feature /note="match: STS: Em:G31644"
repeat_region /note="AluSx repeat: matches 4. .305 of consensus"
misc_feature complement(15538. .15933)
repeat_region /note="match: STS: Em:HSP50B05"
repeat_region /note="MIR repeat: matches 1. .262 of consensus"
repeat_region /note="L1MB4 repeat: matches 3896. .3961 of consensus"
repeat_region /note="AluSg repeat: matches 1. .305 of consensus"
repeat_region /note="AluSp repeat: matches 19. .293 of consensus"
repeat_region /note="L2 repeat: matches 2688. .2746 of consensus"
repeat_region /note="L1PA5 repeat: matches 13. .2993 of consensus"
repeat_region /note="AluY repeat: matches 1. .302 of consensus"
repeat_region /note="L1PA5 repeat: matches 2993. .6141 of consensus"
repeat_region /note="AluSg repeat: matches 1. .310 of consensus"
```

```
repeat_region /note="Charlie5 repeat: matches 4. .116 of consensus"
repeat_region /note="Charlie5 repeat: matches 2241. .2391 of consensus"
repeat_region /note="L1MB4 repeat: matches 5628. .5767 of consensus"
repeat_region /note="THEIC repeat: matches 1. .371 of consensus"
repeat_region /note="THEIC-internal repeat: matches 500. .1580 of consensus"
repeat_region /note="AluSg repeat: matches 1. .293 of consensus"
repeat_region /note="THEIC-internal repeat: matches 3. .500 of consensus"
repeat_region /note="AluY repeat: matches 1. .311 of consensus"
repeat_region /note="THEIC repeat: matches 1. .371 of consensus"
repeat_region /note="L1MB4 repeat: matches 5764. .6182 of consensus"
repeat_region /note="THE1-INTERNAL repeat: matches 752. .877 of consensus"
misc_feature /note="match: GSS: Em:A2576617"
repeat_region /note="85 copies 2 mer tt 67% conserved"
repeat_region /note="3 copies 49 mer 78% conserved"
repeat_region /note="AluSx repeat: matches 1. .309 of consensus"
repeat_region /note="AluSg repeat: matches 1. .307 of consensus"
repeat_region /note="FLAM_A repeat: matches 1. .133 of consensus"
repeat_region /note="AluSg repeat: matches 1. .308 of consensus"
repeat_region /note="MIR repeat: matches 71. .209 of consensus"
repeat_region /note="L2 repeat: matches 2640. .2749 of consensus"
misc_feature /note="match: GSS: Em:A0106154"
repeat_region /note="AluSg/x repeat: matches 128. .290 of consensus"
misc_feature complement(13683. .14100)
misc_feature /note="match: GSS: Em:AQ265669"
repeat_region /note="match: GSS: Em:A0683800"
misc_feature /note="match: STS: Em:G31644"
repeat_region /note="AluSx repeat: matches 4. .305 of consensus"
misc_feature complement(15538. .15933)
repeat_region /note="match: STS: Em:HSP50B05"
repeat_region /note="MIR repeat: matches 1. .262 of consensus"
repeat_region /note="L1MB4 repeat: matches 3896. .3961 of consensus"
repeat_region /note="AluSg repeat: matches 1. .305 of consensus"
repeat_region /note="AluSp repeat: matches 19. .293 of consensus"
repeat_region /note="L2 repeat: matches 2688. .2746 of consensus"
repeat_region /note="L1PA5 repeat: matches 13. .2993 of consensus"
repeat_region /note="AluY repeat: matches 1. .302 of consensus"
repeat_region /note="L1PA5 repeat: matches 2993. .6141 of consensus"
repeat_region /note="AluSg repeat: matches 1. .310 of consensus"
```

Query Match 95.2%; Score 477; DB 9; Length 94727;
Best Local Similarity 97.0%; Pred. No. 3,6e-106;
Matches 486; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
QY 1 CTTTGGTGTGTTTACATGAAGTCCTTGCCCATGCTCTGATGCTGATGCTGCTG 60

```
Db 37088 CTTTGGTGTATTAGACATGAAGTCCTTGCCCATGCCTATGCTCCTGAATGCTATGGCTA 37147
QY 61 GGTTCCTCTAGGAGATTTATGTTTATGCTCTAAGCTTAAGCTTTTAATCCATCTTG 120
Db 37148 GGTTCCTCTAGGAGATTTATGTTTATGCTCTAAGCTTTAATCCATCTTG 37207
QY 121 AATTAATTTTGTATAGGTGTAAGGAAGGATCCAGTTTCAGCTTTCTACATAGGGCTA 180
Db 37208 AATTAATTTTGTATAGGTGTAAGGAAGGATCCAGTTTCAGCTTTCTACATAGGGCTA 37267
QY 181 GCCAGTTTTCAGACACATTTAATAAGGGAATCCTTCCCATGCTGTTTCT 240
Db 37268 GCCAGTTTTCAGACACATTTAATAAGGGAATCCTTCCCATGCTGTTTCT 37327
QY 241 CAGGTTTCTCAAGATCAGATAGTTGTAGATATCGGCATTTATTTCTGAGGCTCTGTC 300
Db 37328 CAGGTTTCTCAAGATCAGATAGTTGTAGATATCGGCATTTATTTCTGAGGCTCTGTC 37387
QY 301 TGTTCATGTTGATATCTCTGTTTGGTACCAGTACCATGTTGTTTGTGTTACTGCTAG 360
Db 37388 TGTTCATGTTGATATCTCTGTTTGGTACCAGTACCATGTTGTTTGTGTTACTGCTAG 37447
QY 361 CTTGTAGTGTAGTTTGAAGTCAGTAGCATGATGCCCTCCAGCTTGTCTTTGGCTTA 420
Db 37448 CTTGTAGTGTAGTTTGAAGTCAGTAGCATGATGCCCTCCAGCTTGTCTTTGGCTTA 37507
QY 421 GGATTGACTTGGCATGCGGCTCTTTTGGTCCACATCAACTTTAAAGTACTGTTT 480
Db 37508 GGATTGACTTGGCATGCGGCTCTTTTGGTCCACATCAACTTTAAAGTACTGTTT 37567
QY 481 CCAATTCGTGAAGAAAGTAA 501
Db 37568 CCAATTCGTGAAGAAAGTCA 37588
```

```
RESULT 8
AC044782/c
LOCUS AC044782 149041 bp DNA linear PRI 23-APR-2002
DEFINITION Homo sapiens chromosome 10 clone RP11-170M17, complete sequence.
ACCESSION AC044782
VERSION AC044782.6 GI:19774346
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 149041)
Smith,D.R.
Genome Therapeutics Corporation Sequencing Center: Human Genome
Sequence Data
Unpublished
2 (bases 1 to 149041)
Smith,D.R.
Direct Submission
Submitted (12-APR-2000) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
3 (bases 1 to 149041)
Smith,D.R.
Direct Submission
Submitted (24-JUL-2001) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
4 (bases 1 to 149041)
Smith,D.R.
Direct Submission
Submitted (28-MAR-2002) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
5 (bases 1 to 149041)
Smith,D.R.
Direct Submission
Submitted (23-APR-2002) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
On Mar 26, 2002 this sequence version replaced gi:15004887.
```

```
FEATURES
source
Location/Qualifiers
1..149041
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-170M17"
/clone_lib="RPC1-11"
BASE COUNT 46956 a 31551 c 29426 g 41108 t
ORIGIN
Query Match 95.28; Score 477; DB 9; Length 149041;
Best Local Similarity 97.08; Pred. No. 3.4e-106;
Matches 486; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
QY 1 CTTTGGTGTATTAGACATGAAGTCCTTGCCCATGCCTATGCTCCTGAATGCTATGGCTG 60
Db 53820 CTTTGGTGTATTAGACATGAAGTCCTTGCCCATGCCTATGCTCCTGAATGCTATGGCTA 53761
QY 61 GGTTCCTCTAGGAGATTTATGTTTATGCTCTAAGCTTTAAGCTTTTAATCCATCTTG 120
Db 53760 GGTTCCTCTAGGAGATTTATGTTTATGCTCTAAGCTTTAAGCTTTTAATCCATCTTG 53701
QY 121 AATTAATTTTGTATAGGTGTAAGGAAGGATCCAGTTTCAGCTTTCTACATAGGGCTA 180
Db 53700 AATTAATTTTGTATAGGTGTAAGGAAGGATCCAGTTTCAGCTTTCTACATAGGGCTA 53641
QY 181 GCCAGTTTTCAGACACATTTAATAAGGGAATCCTTCCCATGCTGTTTCT 240
Db 53640 GCCAGTTTTCAGACACATTTAATAAGGGAATCCTTCCCATGCTGTTTCT 53581
QY 241 CAGGTTTCTCAAGATCAGATAGTTGTAGATATCGGCATTTATTTCTGAGGCTCTGTC 300
Db 53580 CAGGTTTCTCAAGATCAGATAGTTGTAGATATCGGCATTTATTTCTGAGGCTCTGTC 53521
QY 301 TGTTCATGTTGATATCTCTGTTTGGTACCAGTACCATGTTGTTTGGTACTGCTAG 360
Db 53520 TGTTCATGTTGATATCTCTGTTTGGTACCAGTACCATGTTGTTTGGTACTGCTAG 53461
QY 361 CTTGTAGTGTAGTTTGAAGTCAGTAGCATGATGCCCTCCAGCTTGTCTTTGGCTTA 420
Db 53460 CTTGTAGTGTAGTTTGAAGTCAGTAGCATGATGCCCTCCAGCTTGTCTTTGGCTTA 53401
QY 421 GGATTGACTTGGCATGCGGCTCTTTTGGTCCACATCAACTTTAAAGTACTGTTT 480
Db 53400 GGATTGACTTGGCATGCGGCTCTTTTGGTCCACATCAACTTTAAAGTACTGTTT 53341
QY 481 CCAATTCGTGAAGAAAGTAA 501
Db 53340 CCAATTCGTGAAGAAAGTCA 53320
```

```
RESULT 9
AC023273/c
LOCUS AC023273 154076 bp DNA linear HTG 21-FEB-2001
DEFINITION Homo sapiens chromosome 2 clone RP11-199D16, WORKING DRAFT
SEQUENCE, 19 unordered pieces.
ACCESSION AC023273
VERSION AC023273.6 GI:13027570
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 154076)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 154076)
Waterston,R.H.
Direct Submission
Submitted (10-FEB-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
```


COMMENT

On Feb 21, 2001 this sequence version replaced gi:11527490.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
----- Project Information -----
Center project name: H_NH0199D16
----- Summary Statistics -----
Sequencing vector: M13, 85%
Sequencing vector: plasmid, 8%
Chemistry: Dye-terminator ET; 85% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 138250 bases at least Q40
Consensus quality: 143237 bases at least Q30
Consensus quality: 146644 bases at least Q20
Insert size: 142000; agarose-fp
Insert size: 152276; sum-of-contigs
Quality coverage: 4.28 in Q20 bases; agarose-fp
Quality coverage: 3.94 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1706: contig of 1706 bp in length
1707 1806: gap of unknown length
1807 3314: contig of 1508 bp in length
3315 3415: gap of unknown length
3416 5225: contig of 1811 bp in length
5226 5325: gap of unknown length
5326 6707: contig of 1382 bp in length
6708 6807: gap of unknown length
6808 8647: contig of 1840 bp in length
8648 8747: gap of unknown length
8748 11645: contig of 2898 bp in length
11646 11745: gap of unknown length
11746 16416: contig of 4671 bp in length
16417 16517: gap of unknown length
16518 22032: contig of 5516 bp in length
22033 26662: gap of unknown length
26663 26762: contig of 4530 bp in length
26763 31641: contig of 4879 bp in length
31642 31741: gap of unknown length
31742 37785: contig of 6043 bp in length
37786 37885: gap of unknown length
37886 46462: contig of 8577 bp in length
46463 46562: gap of unknown length
46563 56511: gap of unknown length
56512 68341: contig of 11831 bp in length
68342 82114: gap of unknown length
82115 82215: contig of 13673 bp in length
82216 96726: gap of unknown length
96727 96827: contig of 14512 bp in length
96828 113874: gap of unknown length
113875 113975: contig of 17048 bp in length
113976 131413: gap of unknown length
131414 131513: contig of 17438 bp in length
131514 154076: gap of unknown length
154077 22564: contig of 22564 bp in length.

FEATURES

source

Location/Qualifiers
1. 154076
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/clone="RP11-199D16"

misc_feature 1. 1706
/note="assembly_name:Contig33"
misc_feature 1807. 3314
/note="assembly_name:Contig34"
misc_feature 3415. 5225
/note="assembly_name:Contig35"
misc_feature 5326. 6707
/note="assembly_name:Contig36"
misc_feature 6808. 8647
/note="assembly_name:Contig37"
misc_feature 8748. 11645
/note="assembly_name:Contig38"
misc_feature 11746. 16416
/note="assembly_name:Contig39"
vector_side:left*
misc_feature 16517. 22032
/note="assembly_name:Contig40"
misc_feature 22133. 26662
/note="assembly_name:Contig41"
misc_feature 26763. 31641
/note="assembly_name:Contig42"
misc_feature 31742. 37784
/note="assembly_name:Contig43"
misc_feature 37885. 46461
/note="assembly_name:Contig44"
misc_feature 46562. 56410
/note="assembly_name:Contig45"
misc_feature 56511. 68341
/note="assembly_name:Contig46"
misc_feature 68442. 82114
/note="assembly_name:Contig47"
misc_feature 82215. 96726
/note="assembly_name:Contig48"
misc_feature 96827. 113874
/note="assembly_name:Contig49"
misc_feature 113975. 131412
/note="assembly_name:Contig50"
misc_feature 131513. 154076
/note="assembly_name:Contig51"

BASE COUNT 44022 a 33471 c 33238 g 41531 t 1814 others
ORIGIN

Query Match 95.2%; Score 477; DB 2; Length 154076;
Best Local Similarity 97.0%; Pred. No. 3.4e-106;
Matches 486; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
Qy 1 CTTTGTGTTTGTAGACATGAAGTCCCTGGCCATGCCTATGCTGAATGTTATGCTG 60
Db 9271 CTTTGTGTTTGTAGACATGAAGTCCCTGGCCATGCCTATGCTGAATGTTATGCTG 9212
Qy 61 GGTTCCTCTAGGATTTTATGGTTTGTAGTCTAAGCATTTAAGTCTTTATCCATCTG 120
Db 9211 GGTTCCTCTAGGATTTTATGGTTTGTAGTCTAAGCATTTAAGTCTTTATCCATCTG 9152
Qy 121 AATTAATTTTGTATAAGGTGTAAGGAAGGATCCAGTTTCAGCTTTCTACATAGGGCTA 180
Db 9151 AATTAATTTTGTATAAGGTGTAAGGAAGGATCCAGTTTCAGCTTTCTACATAGGGCTA 9092
Qy 181 GCCAGTTTCTACAGCACCATTTATTAATAGGGAATCCTTTCCCATGCTGTTTCT 240
Db 9091 GCCAGTTTCTACAGCACCATTTATTAATAGGGAATCCTTTCCCATGCTGTTTCTG 9032
Qy 241 CAGGTTTGTCAAGATCAGATAGTTGTAGATATCGGCATATTTCTCAGGGCTCTGTC 300
Db 9031 CAGGTTTGTCAAGATCAGATAGTTGTAGATATCGGCATATTTCTCAGGGCTCTGTC 8972
Qy 301 TGTTCATTTGTTTCATATCTCTGTTTGTGGTACCATGACCATGTTGTTTGTACTCTAG 360
Db 8971 TGTTCATTTGTTTCATATCTCTGTTTGTGGTACCATGACCATGTTTGTACTCTAG 8912
Qy 361 CCTTGTAGTGTAGTTTGAAGTCAGGTAGCATGATGATGATGATGATGATGATGATGATG 420
Db 361 CCTTGTAGTGTAGTTTGAAGTCAGGTAGCATGATGATGATGATGATGATGATGATGATG

JOURNAL

Submitted (30-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Sep 23, 2000 this sequence version replaced gi:8569120.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

COMMENT

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence.submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L10069
 Center clone name: 250-I-8
 ----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 157936 bases at least Q40
 Consensus quality: 165044 bases at least Q30
 Consensus quality: 167707 bases at least Q20
 Insert size: 182000; agarose-fp
 Insert size: 169090; sum-of-contigs
 Quality coverage: 4.0 in Q20 bases; agarose-fp
 Quality coverage: 4.3 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 25 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 2796: contig of 2796 bp in length
 * 2797 2896: gap of 100 bp
 * 2897 4248: contig of 1352 bp in length
 * 4249 4348: gap of 100 bp
 * 4349 5349: contig of 1001 bp in length
 * 5350 5449: gap of 100 bp
 * 5450 7045: contig of 1596 bp in length
 * 7046 7145: gap of 100 bp
 * 7146 8600: contig of 1455 bp in length
 * 8601 8700: gap of 100 bp
 * 8701 11899: contig of 3199 bp in length
 * 11900 11999: gap of 100 bp
 * 12000 15340: contig of 3341 bp in length
 * 15341 15440: gap of 100 bp
 * 15441 17756: contig of 2316 bp in length
 * 17757 17856: gap of 100 bp
 * 17857 22299: contig of 4443 bp in length
 * 22300 22399: gap of 100 bp
 * 22400 25462: contig of 3063 bp in length
 * 25463 25562: gap of 100 bp
 * 25563 28942: contig of 3380 bp in length
 * 28943 29042: gap of 100 bp
 * 29043 34484: contig of 5442 bp in length
 * 34485 34584: gap of 100 bp
 * 34585 38797: contig of 4213 bp in length
 * 38798 38897: gap of 100 bp
 * 38898 45155: contig of 6258 bp in length
 * 45156 45255: gap of 100 bp
 * 45256 50138: contig of 4883 bp in length
 * 50139 50238: gap of 100 bp
 * 50239 57081: contig of 6843 bp in length
 * 57082 57181: gap of 100 bp
 * 57182 62192: contig of 5011 bp in length
 * 62193 62292: gap of 100 bp
 * 62293 68432: contig of 6140 bp in length
 * 68433 68532: gap of 100 bp
 * 68533 78869: contig of 10337 bp in length
 * 78870 78969: gap of 100 bp
 * 78970 85886: contig of 6917 bp in length

* 85887 85986: gap of 100 bp
 * 85987 98968: contig of 12982 bp in length
 * 98969 99068: gap of 100 bp
 * 99069 111636: contig of 12568 bp in length
 * 111637 111736: gap of 100 bp
 * 111737 129140: contig of 17404 bp in length
 * 129141 129240: gap of 100 bp
 * 129241 152513: contig of 23273 bp in length
 * 152514 152613: gap of 100 bp
 * 152614 171490: contig of 18877 bp in length.

FEATURES

Source

1. 171490
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="2"
 /clone="RP11-250I8"
 /clone_lib="RPC1-11 Human Male BAC"
 1. 2796
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:left
 2897. 4248
 /note="assembly_fragment"
 4349. 5349
 /note="assembly_fragment"
 5450. 7045
 /note="assembly_fragment"
 7146. 8600
 /note="assembly_fragment"
 8701. 11899
 /note="assembly_fragment"
 12000. 15340
 /note="assembly_fragment"
 15441. 17756
 /note="assembly_fragment"
 17857. 22299
 /note="assembly_fragment"
 22400. 25462
 /note="assembly_fragment"
 25563. 28942
 /note="assembly_fragment"
 29043. 34484
 /note="assembly_fragment"
 34585. 38797
 /note="assembly_fragment"
 38898. 45155
 /note="assembly_fragment"
 45256. 50138
 /note="assembly_fragment"
 50239. 57081
 /note="assembly_fragment"
 57182. 62192
 /note="assembly_fragment"
 62293. 68432
 /note="assembly_fragment"
 68533. 78869
 /note="assembly_fragment"
 78970. 85886
 /note="assembly_fragment"
 85987. 98968
 /note="assembly_fragment"
 99069. 111636
 /note="assembly_fragment"
 111737. 129140
 /note="assembly_fragment"
 129241. 152513
 /note="assembly_fragment"
 152614. 171490
 /note="assembly_fragment"
 . clone_end:T7
 vector_side:right

BASE COUNT 50305 a 35432 c 33845 g 49505 t 2403 others

Query Match 95.2%; Score 477; DB 2; Length 176630;
Best Local Similarity 97.0%; Pred. No. 3.4e-106;
Matches 486; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 CTTTGGTGTAGACATGAAGTCCCTTGCCCATGCCCTATGCTCTGAATGGTATGCGCTG 60
DB CTTTGGTGTGTAGACATGAAGTCCCTTGCCCATGCCCTATGCTCTGAATGGTATGCGCTA 7151

QY 61 GGTTCCTCTCAGGATTTTATGCTTTAGTCTAAACATTTAAGTCTTTAATCCATCTTG 120
DB GGTTCCTCTCAGGATTTTATGCTTTAGTCTAAACATTTAAGTCTTTAATCCATCTTG 7091

QY 121 AATTAATTTTGTATAAGGTGTAAGGAAGGATCCAGTTTCTACATAGGCGTA 180
DB AATTAATTTTGTATAAGGTGTAAGGAAGGATCCAGTTTCTACATAGGCGTA 7031

QY 181 GCCAGTTTCTCAGCACCATTTATTAATAGGGAATCCTTCCCATGCTGTTTTCT 240
DB GCCAGTTTCTCAGCACCATTTATTAATAGGGAATCCTTCCCATGCTGTTTTCT 6971

QY 241 CAGGTTTGTCAAGATCAGATAGTTAGATATGCGGATTTATTTCTGAGGGCTCTGTC 300
DB CAGGTTTGTCAAGATCAGATAGTTAGATATGCGGATTTATTTCTGAGGGCTCTGTC 6911

QY 301 TGTTCATTTGTTGATATCTCTGTTTGGTACCAGTACCATGTTGTTTGGTTACTGTAG 360
DB TGTTCATTTGTTGATATCTCTGTTTGGTACCAGTACCATGTTGTTTGGTTACTGTAG 6851

QY 361 CCTTGTAGTGTAGTTGAAGTCAGTACGATGATGCTCCAGCTTGTCTTTTGGCGTTA 420
DB CCTTGTAGTGTAGTTGAAGTCAGTACGATGATGCTCCAGCTTGTCTTTTGGCGTTA 6791

QY 421 GGATTTGCTGGCATGTGGGCTCTTTTGGTTCCACATGAACCTTTAAAGTACTTTTT 480
DB GGATTTGCTGGCATGTGGGCTCTTTTGGTTCCACATGAACCTTTAAAGTACTTTTT 6731

QY 481 CCAATTCGTGAAGAAAGTAA 501
DB CCAATTCGTGAAGAAAGTCA 6710

RESULT 14
AC069543 179789 bp DNA linear PRI 28-MAR-2002
LOCUS Homo sapiens chromosome 10 clone RP11-393H5, complete sequence.
AC069543
AC069543.5 GI:19774362
VERSION HTG.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 179789)
AUTHORS Smith,D.R.
TITLE Genome Therapeutics Corporation Sequencing Center: Human Genome Sequence Data
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 179789)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-2000) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02453, USA
REFERENCE 3 (bases 1 to 179789)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2001) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02453, USA
REFERENCE 4 (bases 1 to 179789)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (28-MAR-2002) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02453, USA

COMMENT On Mar 28, 2002 this sequence version replaced gi:14010785.
FEATURES
Location/Qualifiers
source 1..179789
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-393H5"
/clone_lib="RPC1-11"
BASE COUNT 50198 a 34202 c 37287 g 58102 t
ORIGIN

Query Match 95.2%; Score 477; DB 9; Length 179789;
Best Local Similarity 97.0%; Pred. No. 3.4e-106;
Matches 486; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 CTTTGGTGTGTAGACATGAAGTCCCTTGCCCATGCCCTATGCTCTGAATGGTATGCGCTG 60
DB CTTTGGTGTGTAGACATGAAGTCCCTTGCCCATGCCCTATGCTCTGAATGGTATGCGCTA 153506

QY 61 GGTTCCTCTCAGGATTTTATGCTTTAGTCTAAACATTTAAGTCTTTAATCCATCTTG 120
DB GGTTCCTCTCAGGATTTTATGCTTTAGTCTAAACATTTAAGTCTTTAATCCATCTTG 153566

QY 121 AATTAATTTTGTATAAGGTGTAAGGAAGGATCCAGTTTCTACATAGGCGTA 180
DB AATTAATTTTGTATAAGGTGTAAGGAAGGATCCAGTTTCTACATAGGCGTA 153626

QY 181 GCCAGTTTCTCAGCACCATTTATTAATAGGGAATCCTTCCCATGCTGTTTTCT 240
DB GCCAGTTTCTCAGCACCATTTATTAATAGGGAATCCTTCCCATGCTGTTTTCT 153686

QY 241 CAGGTTTGTCAAGATCAGATAGTTAGATATGCGGATTTATTTCTGAGGGCTCTGTC 300
DB CAGGTTTGTCAAGATCAGATAGTTAGATATGCGGATTTATTTCTGAGGGCTCTGTC 153746

QY 301 TGTTCATTTGTTGATATCTCTGTTTGGTACCAGTACCATGTTGTTTGGTTACTGTAG 360
DB TGTTCATTTGTTGATATCTCTGTTTGGTACCAGTACCATGTTGTTTGGTTACTGTAG 153806

QY 361 CCTTGTAGTGTAGTTGAAGTCAGTACGATGATGCTCCAGCTTGTCTTTTGGCGTTA 420
DB CCTTGTAGTGTAGTTGAAGTCAGTACGATGATGCTCCAGCTTGTCTTTTGGCGTTA 153866

QY 421 GGATTTGCTGGCATGTGGGCTCTTTTGGTTCCACATGAACCTTTAAAGTACTTTTT 480
DB GGATTTGCTGGCATGTGGGCTCTTTTGGTTCCACATGAACCTTTAAAGTACTTTTT 153926

QY 481 CCAATTCGTGAAGAAAGTAA 501
DB CCAATTCGTGAAGAAAGTCA 154006

RESULT 15
AC087774 184000 bp DNA linear PRI 07-MAR-2002
LOCUS Homo sapiens 12p BAC RP11-633013 (Roswell Park Cancer Institute Human BAC Library) complete sequence.
AC087774
AC087774.27 GI:16572889
VERSION HTG.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 184000)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbara,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,

Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Emerling, S., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gili, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Han, J., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hoques, M., Holloway, C., Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Ioshikhes, I., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, D., Korvah, J., Kovar, C., Kratovic, J., Kuresh, A., Landry, N., Leal, B., Lee, E., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, J., Liu, J., Liu, W., Loulsegh, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapa, P., Marondel, I., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Merscher, S., Metzker, M., Miller, A., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Montgomery, K.T., Morgan, M., Morris, S., Moser, M., Neal, D., Nelson, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shim, C., Shooshitari, N., Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Kuchelapatti, R., Weinstock, G. and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 184000)
Worley, K.C.
Direct Submission
Submitted (24-JAN-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 184000)
Worley, K.C.
Direct Submission
Submitted (01-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 184000)
Worley, K.C.
Direct Submission
Submitted (02-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 184000)
Worley, K.C.
Direct Submission
Submitted (26-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
6 (bases 1 to 184000)
Worley, K.C.
Direct Submission
Submitted (07-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Nov 1, 2001 this sequence version replaced gi:16327961.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones.

Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

FEATURES	Source	Location/Qualifiers
misc_feature	1. .184000	/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/chromosomes="12"
		/clone="RP11-633013"
	1. .2009	/function="clone overlap"
repeat_region	26. .46	/rpt_family="AT_rich"
repeat_region	1466. .1804	/rpt_family="LIP3"
repeat_region	2274. .2583	/rpt_family="AluSp"
misc_feature	2584	/note="(A or *)"
		/function="unsure base"
repeat_region	complement(2851. .2921)	/rpt_family="LIP"
repeat_region	2922. .5243	/rpt_family="LIP3"
repeat_region	complement(5470. .6019)	/rpt_family="MLTID"
repeat_region	6640. .6756	/rpt_family="HERVL"
repeat_region	7042. .7074	/rpt_family="(TTTA)n"
repeat_region	complement(7092. .7265)	/rpt_family="LIHS"
repeat_region	7563. .8070	/rpt_family="LTRA0a"
repeat_region	8174. .8253	/rpt_family="MERVL"
repeat_region	8254. .8315	/rpt_family="(TA)n"
repeat_region	8316. .8541	/rpt_family="MERVL"
repeat_region	8791. .9287	/rpt_family="MLTIC"
repeat_region	9497. .9538	/rpt_family="AT_rich"

```

Db      71085 CCAATTCTGTGAGAAAGTCA 71105
|||||
Search completed: July 4, 2003, 16:40:47
Job time : 1513.67 secs

```

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 14, 2003, 15:09:32 ; Search time 1093.54 seconds
(without alignments)
7405.060 Million cell updates/sec

Title: US-10-083-853b-2_COPY_1_500

Perfect score: 500

Sequence: 1 gtagatgaagaagcctca.....actaaagttcttcaggct 500

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database

EST

2: em_estb:*

3: em_estm:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_estl:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: gb_gss:*

18: em_gss_hum:*

19: em_gss_inv:*

20: em_gss_pln:*

21: em_gss_vrt:*

22: em_gss_fun:*

23: em_gss_mam:*

24: em_gss_mus:*

25: em_gss_othr:*

26: em_gss_pro:*

27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	500	100.0	606	14	BQ575044
2	500	100.0	3739	11	AF113216
3	471.8	94.4	505	10	AW274723
4	460	92.0	485	9	AA488996
5	370.8	74.2	406	9	AI221902
6	250.8	50.2	467	13	BM144799

C	7	244.4	48.9	560	10	AW968905	
8	202.4	40.5	217	9	AI570720	tr66a05.x	
9	154.6	30.9	318	9	AA904948	o78ld10.s	
10	149.4	29.9	317	9	AA736766	aa23a03.s	
C	11	131.8	26.4	360	9	AA488780	aa54c10.r
C	12	108.6	21.7	1796	11	AK003359	Mus muscu
13	69.4	13.9	924	17	CNS0206S	Tetraodon	
C	14	66.6	13.3	1124	17	CNS073BM	clone BA0
C	15	64.4	12.9	1201	17	CNS016E1	AI106627
C	16	63.2	12.6	1043	17	CNS0145P	Drosophil
C	17	62.6	12.5	1101	17	CNS00EVL	AL103735
C	18	62	12.4	945	17	CNS04D0K	AL069706
C	19	60	12.0	1084	17	CNS071NH	AL285149
C	20	60	12.0	1200	17	CNS016CO	clone BA0
C	21	57.6	11.5	450	13	BI297936	Drosophil
C	22	57.6	11.5	945	17	CNS04D0K	UI-R-CV2-
C	23	57.2	11.4	1101	17	CNS0039G	AL285149
C	24	57	11.4	1101	17	CNS00E07	Tetraodon
C	25	56.6	11.3	1092	17	CNS020K7	AL063921
C	26	56.4	11.3	928	17	CNS00DKY	Drosophil
C	27	56.4	11.3	1101	17	CNS0039G	AL063921
C	28	56.2	11.2	418	13	BI297322	Drosophil
C	29	56.2	11.2	427	13	BI297805	UI-R-CV2-
C	30	56.2	11.2	466	13	BI297955	UI-R-CV2-
C	31	56.2	11.2	482	12	BF406722	UI-R-BJ2-
C	32	56.2	11.2	489	9	AA997136	UI-R-CV1-
C	33	56.2	11.2	566	12	BG374585	UI-R-CV1-
C	34	56.2	11.2	571	13	BI301086	UI-R-CV2-
C	35	56.2	11.2	650	13	BI296665	UI-R-CV2-
C	36	56.2	11.2	758	13	BI296653	UI-R-CV2-
C	37	55.6	11.1	1101	17	CNS0172Q	AL108704
C	38	54.8	11.0	676	17	BH648307	BOHYL18TR
C	39	54.6	10.9	428	13	BI297921	UI-R-CV2-
C	40	54.6	10.9	441	12	BG374397	UI-R-CV1-
C	41	54.4	10.9	1037	14	BQ648574	AGENCOUTF
C	42	54	10.8	1201	17	CNS016E1	AL106627
C	43	53.8	10.8	1013	17	CNS06RPQ	Drosophil
C	44	53.6	10.7	928	17	CNS00DKY	AL071865
C	45	53.4	10.7	1101	17	CNS00EVL	AL069706

ALIGNMENTS

RESULT 1
BQ575044
LOCUS
DEFINITION
BQ575044
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BQ575044
UI-H-EZ1-bbd-o-15-0-UI-s1 NCI-CGAP-Ch2 Homo sapiens cDNA clone
UI-H-EZ1-bbd-o-15-0-UI 3', mRNA sequence.
BQ575044
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 606)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue procurement: Dr. Steven Gitelis/ Rush Presbiterian, Dept. of
Orthopaedics

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bento-soares@uiowa.edu
The following repetitive elements were found in this cDNA
sequence: 1-24, >AT-rich#Low_complexity 122-142,
>AT-rich#Low_complexity 249-325, >(TA)n#Simple_repeat


```

Db      1323 CAGGCTAGTAAGTTGGTGGTAGGAAAAGGCTCTCTTATCTCACCCCTCTTAA 1264
QY      481 ACTAAAGCTTCTTTCAGGCT 500
Db      1263 ACTAAAGCTTCTTTCAGGCT 1244

RESULT 3
AW274723
LOCUS   AW274723
DEFINITION x334b02.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2695563 3',
mRNA sequence.
ACCESSION AW274723.1
VERSION   AW274723.1
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 505)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-r@mail.nih.gov
          Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
          Emmert-Buck, M.D., Ph.D.
          cDNA Library Preparation: M. Bento Soares, Ph.D.
          DNA Sequencing by: Greg Lennon, Ph.D.
          DNA Sequencing by: Washington University Genome Sequencing Center
          Clone distribution: NCI-CGAP clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          www-bio.llnl.gov/bbrp/image/image.html
          Seq primer: -40UP from Gibco
          High quality sequence stop: 475.
          Location/Qualifiers
            1..505
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="IMAGE:2695563"
              /clone_lib="NCI_CGAP_Kid11"
              /lab_host="DH10B"
              /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
              a modified polylinker; Site 1: Not I; Site 2: Eco RI;
              Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
              prepared, and ss circles were made in vitro. Following HAP
              purification, this DNA was used as tracer in a subtractive
              hybridization reaction. The driver was PCR-amplified cDNAs
              from a pool of 5,000 clones made from the same library
              (cloneIDs 132376-132391, 1456007-1456775, and
              1500552-1502855). Subtraction by Bento Soares and M.
              Fatima Bonaldo."
BASE COUNT 169 a 83 c 75 g 178 t
ORIGIN

Query Match 94.4%; Score 471.8; DB 10; Length 505;
Best Local Similarity 99.2%; Pred No. 6.4e-93;
Matches 485; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

QY      1 GTATATGTAAGAACGCCCTCATCTTTTGATTTTAAATATACAGATGCTTCTTTAAAGAGA 60
Db      17 GTATATGTAAGAACGCCCTCATCTTTTGATTTTAAATATACAGATGCTTCTTTAAAGAGA 76
QY      61 GCAGATTCAAATTTGTTTGTTCCTTCAAAATTTTAAATATATCTCTCTTAATTTT 120
Db      77 GCAGATTCAAATTTGTTTGTTCCTTCAAAATTTTAAATATATCTCTCTTAATTTT 136
QY      121 CTAAGACATCTTTCATATATTTGACCATCCCTTATTTTGGCAAGGATTTTAAAGAGTCT 180
Db      137 CTAAGACATCTTTCATATATTTGACCATCCCTTATTTTGGCAAGGATTTTAAAGAGTCT 196
QY      181 AACTCAACATATGTAAGCTCTGGTGTAACCTGGTATATATATACC--AAAAAAAACATTTG 238

```

```

Db      137 AACTCAACATATGTAAGCTCTGGTGTAACCTGGTATATATATACCAAAAAAACAATTTG 256
QY      239 ATCTATATACACATAGACATGAATATATTTCTGTGTGTGTGTGTGCATATATAACCTCAA 298
Db      257 ATCTATATACACATAGACATGAATATATTTCTGTGTGTGTGTGTGCATATATAACCTCAA 316
QY      299 ACACATATTATTAATGCAATCCTATATTTCTTAGGTATAGAAAGTTGATGATATACCTTTCT 358
Db      317 ACACATATTATTAATGCAATCCTATATTTCTTAGGTATAGAAAGTTGATGATATACCTTTCT 376
QY      359 ACTTGCCATGGCATTTAACAAGCAAGGCTGAGACTCAGCAACCACTTGTGTTCATTGCAT 418
Db      377 ACTTGCCATGGCATTTAACAAGCAAGGCTGAGACTCAGCAACCACTTGTGTTCATTGCAT 436
QY      419 TGCAGGCTAGTAGTAAGTTTGGTTCCTGGTAGGAAAAGGCTCTCTTATCTCACCCCTCTT 478
Db      437 TGCAGGCTAGTAGTAAGTTTGGTTCCTGGTAGGAAAAGGCTCTCTTATCTCACCCCTCTT 496
QY      479 AAACATAAAG 487
Db      497 AAACATAAAG 505

RESULT 4
AA488996
LOCUS   aa54c10.s1 NCI_CGAP_GCBI Homo sapiens cDNA clone IMAGE:824754 3',
mRNA sequence.
ACCESSION AA488996
VERSION   AA488996.1
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 485)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-r@mail.nih.gov
          Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
          Ph.D., Gerald Marti, M.D.
          cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
          Bonaldo, Ph.D.
          cDNA Library Arrayed by: Greg Lennon, Ph.D.
          DNA Sequencing by: Washington University Genome Sequencing Center
          Clone distribution: NCI-CGAP clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          www-bio.llnl.gov/bbrp/image/image.html
          Seq primer: -4lm13 fwd. Et from Amersham
          High quality sequence stop: 465.
          Location/Qualifiers
            1..485
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="IMAGE:824754"
              /clone_lib="NCI_CGAP_GCBI"
              /tissue_type="germinal center B cell"
              /lab_host="DH10B"
              /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
              polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
              was prepared from human tonsillar cells enriched for
              germinal center B cells by flow sorting (CD20+, IgD-),
              provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
              (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
              primed with a Not I - oligo(dT) primer
              15'-TGTTACCAATCTGAAGTGGAGCGCGCGCTCATTTTTTTTTTTTTTTT-3'
              1. Double-stranded cDNA was ligated to Eco RI adaptors
              (Pharmacia), digested with Not I and cloned into the Not I
              - and Eco RI sites of the modified pT7T3 vector. Library

```


Seq primer: M13 primer.
Location/Qualifiers
1. .467
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="TCAAP1310"
/clone_lib="Pediatric acute myelogenous leukemia cell (FAB M1) Baylor-HGSC project-TCAA"
/sex="male"
/tissue_type="leukopheresis"
/cell_type="myeloid cell"
/dev_stage="pediatric 6 years"
/lab_host="DH10B"
/note="vector: lambda PSB; Site_1: BamHI; Site_2: EcoRI; First strand cDNA was primed with an anchored xhoI-oligo(dT) primer [5'GGAGGACTGCGCGCCGAGGAG(T)VN 3'; V-A,C,G; N-A,C,G,T] and then dG tailed. Second strand was primed with a BamHI-dC primer [5'AGAGAGTCGATCCGCGCGCCCAATAATAAT(C) 3']. Double-stranded cDNA was then digested with BamHI and XhoI and directionally cloned into the BamHI and SalI sites of lambda PSB vector. Library went through one round of normalization. Library was constructed by Wei Yu at RIKEN of Japan (Carninci P, Westover A, Nishiyama Y, Ohsumi T, Itoh M, Nagaoka S, Sasaki N, Okazaki Y, Muramatsu M, Schneider C, Hayashizaki Y, High efficiency selection of full-length cDNA by improved biotinylated cap trapper., DNA Res 4: 1, 61-6, Feb 28, 1997)."
BASE COUNT 159 a | 65 c. 180 t | 1 others
ORIGIN
Query Match 50.2%; Score 250.8; DB 13; Length 467;
Best Local Similarity 97.8%; Pred. No. 8.7e-45;
Matches 266; Conservative 0; Mismatches 2; Indels 4; Gaps 1;
Qy 1 GTATATGTAGAAGCGCTCACTTTTGGATTTTAAATATACAAGATGCTTTCTTTAAGAGA 60
Db 195 GTATATGTAGAAGCGCTCACTTTTGGATTTTAAATATACAAGATGCTTTCTTTAAGAGA 254
Qy 61 GCAAGATTCAAAATGTTTGTGTTTCAAAATTTAAATATATATCTCTCTAAATTTT 120
Db 255 GCAAGATTCAAAATGTTTGTGTTTCAAAATTTAAATATATATCTCTCTAAATTTT 314
Qy 121 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAGAGATTTTAAGAGTCT 180
Db 315 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAGAGATTTTAAGAGTCT 374
Qy 181 AACTCAACATATGTAGCTCTGTGTACCTGGTTATATATACC----AAAAAACAATT 236
Db 375 AACTCAACATATGTAACTCTGTGTACCTGGTTATATATATACCAAAAAAACAATT 434
Qy 237 TGATCTATATACATAGATGAATATATTT 268
Db 435 TGAATCTATATACATAGATGAATATATTT 466
RESULT 7
AW968905/c
LOCUS AW968905 560 bp mRNA linear EST 01-JUN-2000
DEFINITION EST380981 MAGE resequences, MAGJ Homo sapiens cDNA, mRNA sequence.
ACCESSION AW968905
VERSION AW968905.1 GI:8158746
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Hegde,P., Oi,R., Abernathy,K., Dharap,S., Gaspar,D., Gay,C., Holt I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and Quackenbush,J.
TITLE Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray

Unpublished (2000)
Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Plate: 258
Seq primer: Forward.
Location/Qualifiers
1. .560
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGJ"
/note="vector: pbluescriptSKm"
BASE COUNT 163 a | 122 c | 120 g | 155 t
ORIGIN
Query Match 48.9%; Score 244.4; DB 10; Length 560;
Best Local Similarity 92.4%; Pred. No. 2.1e-43;
Matches 268; Conservative 0; Mismatches 21; Indels 1; Gaps 1;
Qy 212 GGTATATATATACCAAAAAACATTTGATCTATATACATAGACATGAATATAT-TTCT 270
Db 558 GGTGACATATAGGAAGCAAAACATTTGATCTAGATCTATAGAGCGTGGATATATAT 499
Qy 271 GTGTGTGTTTGTGCATATATAACCTCAAAACACTATTATTAATGCAATCCTATATCTTA 330
Db 498 GTGTGTGTTTGGCAAGTATAACCTCAAGCACTATAAGTAATGCAATCCTATATCTTA 439
Qy 331 GGTATAGAAGTTGATATATACCTTTTCTTACTTGGCATGGCATTAACAAAGCAAGCTGAG 390
Db 438 GGTATAGAAGTTGATATATAACCTTTTGTACTTGGCATGGCATTAACAAAGCAAGCTGAG 379
Qy 391 ACTCAGCAACCACTGTGTTTCATTCGATTCAGGCTAGTAGTAGTTGGTTGCTGGTAG 450
Db 378 ACTCAGCAACCACTGTGTTTCATTCGATTCAGGCTAGTAGTAGTTGGTTGCTGGTAG 319
Qy 451 GAAAGGGTCTCTTATCTCACCCTCCTTAAAGGTTCTTTCAGGCT 500
Db 318 GAAAGGGTCTCTTATCTCACCCTCCTTAAAGTAAATGTTTCTTTCAGGCT 269
RESULT 8
AI570720 217 bp mRNA linear EST 14-MAY-1999
LOCUS tr56a05.x1 NCI-CGAP_Paul Homo sapiens cDNA clone IMAGE:223248 3',
DEFINITION mRNA sequence.
ACCESSION AI570720
VERSION AI570720.1 GI:4534094
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
CONTACT: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Life Technologies catalog #: 11548-013
DNA Sequencing by: Washington University Genome Sequencing Center
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1407 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 186
POLYA-No. Location/Qualifiers
1. .217

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:223248"
/clone_lib="NCI_CGAP_Pan1"
/tissue_type="adenocarcinoma"
/lab_host="DH108"
/notes="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.72 kb. Life Technologies catalog #:
11548-013"
```

BASE COUNT	77 a	30 c	29 g	81 t
ORIGIN				
Query Match	40.5%; Score 202.4; DB 9; Length 217;			
Best Local Similarity	99.5%; Pred. No. 3.2e-34;			
Matches 203; Conservative	0; Mismatches 1; Indels 0; Gaps 0;			

```
QY 1 GTATATGTAAGAAGCCTCATCTTTTGATTTTAAATATACAGATGCTTCTTTAAGAGA 60
|||||
Db 14 GTATATGTAAGAAGCCTCATCTTTTGATTTTAAATATACAGATGCTTCTTTAAGAGA 73
|||||
QY 61 GCAAGATTCAAAATCTTTTGCTTTCAAAATTTAAATAAATATATCTCTAAATTTT 120
|||||
Db 74 GCAAGATTCAAAATCTTTTGCTTTCAAAATTTAAATAAATATATCTCTAAATTTT 133
|||||
QY 121 CTAAGACATGTTTCATATATTGACCATCCCTTATTGGCAAGGATTTTAAAGAGTCT 180
|||||
Db 134 CTAAGACATGTTTCATATATTGACCATCCCTTATTGGCAAGGATTTTAAAGAGTCT 193
|||||
QY 181 AACTCAACATATGTAAGCTCTGG 204
|||||
Db 194 AACTCAACATATGTAAGCTCTGG 217
|||||
```

RESULT 9
AA904948
LOCUS
DEFINITION
oJ81b10.s1 Soares_NFL_T_GBC.S1 Homo sapiens cDNA clone
IMAGE:1504699 3', mRNA sequence.
AA904948
ACCESSION
VERSION
KEYWORDS
SOURCE
EST.
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 318)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (infoimage.llnl.gov) for further information.
Insert Length: 1030 Std Error: 0.00
Seq primer: -40m13 fwd. Et from Amersham
High quality sequence stop: 152.
Location/Qualifiers
1. .318
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1504699"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH108"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NbHL19w, testis NHT, and B-cell
NCI-CGAP.GC81) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made

from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo.

BASE COUNT	101 a	47 c	47 g	123 t
ORIGIN				
Query Match	30.9%; Score 154.6; DB 9; Length 318;			
Best Local Similarity	97.5%; Pred. No. 8.1e-24;			
Matches 157; Conservative	0; Mismatches 4; Indels 0; Gaps 0;			

```
QY 1 GTATATGTAAGAAGCCTCATCTTTTGATTTTAAATATACAGATGCTTCTTTAAGAGA 60
|||||
Db 158 GTATATGTAAGAAGCCTCATCTTTTGATTTTAAATATACAGATGCTTCTTTAAGAGA 217
|||||
QY 61 GCAAGATTCAAAATCTTTTGCTTTCAAAATTTAAATAAATATATCTCTAAATTTT 120
|||||
Db 218 GCAAGATTCAAAATCTTTTGCTTTCAAAATTTAAATAAATATATCTCTAAATTTT 277
|||||
QY 121 CTAAGACATGTTTCATATATTGACCATCCCTTATTTTGG 161
|||||
Db 278 CTAAGACATGTTTCATATATTGACCATCCCTTATTTTGG 318
|||||
```

RESULT 10
AA736766
LOCUS
DEFINITION
oa23a03.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1305772 3',
mRNA sequence.
AA736766
ACCESSION
VERSION
KEYWORDS
SOURCE
EST.
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 317)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., David Allman,
Ph.D., Gerald Marti, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/[LLNL](http://www-bio.llnl.gov/bbrp/image/image.html) at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1055 Std Error: 0.00
Seq primer: -40m13 fwd. Et from Amersham
High quality sequence stop: 252.
Location/Qualifiers
1. .317
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1305772"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH108"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-GGTACCAATCTGAAGTGGAGCGCGCTCATTTTTTTTTTTT-3'
]. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I


```

source
1. .924
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="157A04"
/clone_lib="G"
/notes="Genoscope sequence ID : COAG157BA02SP1-end :
PUC-Ori"
BASE COUNT 300 a 109 c 172 g 252 t 91 others
ORIGIN
Query Match 13.9%; Score 69.4; DB 17; Length 924;
Best Local Similarity 44.8%; Pred. No. 2.6e-05;
Matches 130; Conservative 28; Mismatches 132; Indels 0; Gaps 0;

QY 70 AAAATGTTTGTGTTTCAAAATTTAAATAATAATTTATCTCTCTAAATTTCTTAAAGACA 129
||||| ||||| ||||| ||||| ||||| :||| : ||||| ||||| |||||
Db 601 AAAATTTTATTTTATTTAAATTTAAATTTAAATTTAAATTTAAATTTTAAATTTTAAATTT 660
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 130 TGTTCATATATTTGACCATCCCTTATTTTGGCAAGAGTTTAAAGAGTCTACTCAAC 189
||| :||| ||||| :||| ||||| :||| ||||| :||| ||||| :||| ||||| :|||
Db 661 TTTTATTTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 720
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 190 ATATGACGCTCTGGTGCTACCTGGTTATATATACCAAAAAACATTTGATCTATATACA 249
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 721 AWAATTAATAAATTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 780
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 250 CATAGACATGAATATATTTCTGTGCTGTTTGGCATATATATAACCTCAACACTATTATT 309
||| :||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 781 AWAATTAATAAATTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 840
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 310 AATGCAATCTATATCTTAGTATAGATAGAGTTGATGATATACCTTTCTA 359
||| :||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 841 AATWAAATTTTATATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 890
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 14
CNS073BM/c
LOCUS
DEFINITION
1124 bp DNA linear GSS 07-JUL-2001
clone BA0AB018D03 of library BA0AB from strain CLIB 210 of
Kluyveromyces lactis, genomic survey sequence.
ACCESSION
AL427304.1 GI:12210498
VERSION
GSS.
KEYWORDS
Kluyveromyces lactis.
SOURCE
Kluyveromyces lactis
ORGANISM
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
REFERENCE
1 (bases 1 to 1124)
Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissenbach,J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
JOURNAL
FEBs Lett. 487 (1), 3-12 (2000)
MEDLINE
20584711
PUBMED
11152876
REFERENCE
2 (bases 1 to 1124)
Bolotin-Fukuhara,M., Toffano-Nioche,C., Artiguenave,F.,
Duchateau-Nguyen,G., Lemaire,M., Marmeisse,R., Montrocher,R.,
Robert,C., Termier,M., Wincker,P. and Wesolowski-Louvel,M.
Genomic exploration of the hemiascomycetous yeasts: 11.
Kluyveromyces lactis
JOURNAL
FEBs Lett. 487 (1), 66-70 (2000)
MEDLINE
20584721
PUBMED
11152886
REFERENCE
3 (bases 1 to 1124)
Genoscope.
Direct Submission
JOURNAL
Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremlieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :

```

```

seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
FEATURES
source
1. .1124
/organism="Kluyveromyces lactis"
/strain="CLIB 210"
/variety="lactis"
/db_xref="taxon:28985"
/clone="BA0AB018D03"
/clone_lib="BA0AB"
BASE COUNT 392 a 96 c 109 g 409 t 118 others
ORIGIN
Query Match 13.3%; Score 66.6; DB 17; Length 1124;
Best Local Similarity 46.8%; Pred. No. 0.00011;
Matches 153; Conservative 19; Mismatches 155; Indels 1; Gaps 1;

QY 9 AAGAAAGCCCTCATCTTTTGAATTTTAAATATACAGATGCTTTTAAAGAGCAAGATT 68
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1009 AAAAAAATAAATAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 950
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 69 CAAAAATGTTTGTGTTTCAAAATTTAAATAATAATTTATCTCTCTAAATTTCTTAAAGAC 128
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 949 AAATATTTTATATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 890
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 129 ATGTTTCATATATTGACCATCCCTTATTTTGCAGAGGATTTTAAGAGTCTACTCAA 188
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 889 ATTNATTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTAT 830
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 189 C-ATATGTAAGCTCTGCTGTACCTGTATATATACCAAAAAACATTTGATCTATATA 247
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 829 TTAAATATTTATTTATTTATTTTATTTTATTTTATTTTATTTTATTTATTTATTTAT 770
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 248 CACATGACATGAATATATTTCTGTGCTGTTTGTGTCATATATAACCTCAACACTATTA 307
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 769 ATNTTTATTTATTTATTTATTTATTTCTNTAAATTTTAAATTTTAAATTTTAAATTTAT 710
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 308 TTAATGCAATCTATATTTCTTAGGTAT 335
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 709 TTAATAATTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 682
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15
CNS016EL/c
LOCUS
DEFINITION
1201 bp DNA linear GSS 26-JUL-1999
Drosophila melanogaster genome survey sequence S96 end of BAC
BACN15C13 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
AL106627
VERSION
AL106627.1 GI:5622848
KEYWORDS
GSS.
SOURCE
Drosophila melanogaster.
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 1201)
Genoscope.
Direct Submission
JOURNAL
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 4, 2003, 19:30:58 ; Search time 146 seconds
(without alignments)
12059.771 Million cell updates/sec

Perfect score: 52528

Sequence: 1 gttatgtaagaagcctca.....caattctgtgaagaagtaa 29921

Scoring table: BLOSUM62

Xgapext 10.0, Xgapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p model -DEV=xlh

-Q=/cgn2_1/USPTO_spool/US10083853/runat_03072003_093612_8118/app_query.fasta_1.30087

-DB=Issued Patents AA -QMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0

-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi

-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15

-MODE=LOCAL -OUTFMT=plc -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000

-USER=US10083853.scgn.1.1.221.runat.03072003_093612_8118 -NCPU=6 -ICPU=3

-NO_MMAP -LARGUEURY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG

-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6

-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

-Database

Issued Patents AA:

1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep.*

2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep.*

3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep.*

4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep.*

5: /cgn2_6/ptodata/1/iaa/PCTUS.COMB.pep.*

6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	409	0.8	370	4	US-09-457-066-37
c 2	409	0.8	370	4	US-09-540-224-2
c 3	382	0.7	370	4	US-09-540-224-4
c 4	314	0.6	1079	3	US-09-058-489-22
5	311	0.6	132	4	US-09-605-785-573
6	306	0.6	375	2	US-08-454-557C-121
7	306	0.6	375	2	US-08-340-426D-121
8	306	0.6	375	2	US-08-450-673C-121
9	280	0.5	397	5	PCT-US95-17111A-121
10	271.5	0.5	110	4	US-09-227-357-193
c 11	272	0.5	1350	4	US-09-245-041-17
c 12	253	0.5	122	4	US-09-227-357-285

c 13	239	0.5	345	4	US-09-040-220D-2
c 14	239	0.5	345	4	US-09-457-066-2
c 15	239	0.5	345	4	US-09-265-686-2
c 16	239	0.5	345	4	US-09-540-224-5
c 17	235	0.4	99	4	US-09-288-143-168
c 18	235	0.4	345	4	US-09-457-066-43
19	209.5	0.4	10182	4	US-09-134-001C-3159
20	198	0.4	90	4	US-09-227-357-171
21	194.5	0.4	56	4	US-09-227-357-577
22	185.5	0.4	1312	2	US-08-592-126-148
23	185.5	0.4	1312	2	US-08-687-080-51
24	184.5	0.4	1350	4	US-09-245-041-17
25	183	0.3	1588	5	PCT-US93-07261-11
26	183	0.3	1663	5	PCT-US93-07261-16
c 27	181.5	0.3	76	4	US-09-605-785-575
c 28	178	0.3	3248	1	US-08-353-700-1
29	178	0.3	3248	5	PCT-US95-16216-1
30	177	0.3	2482	1	US-08-328-254-6
31	173	0.3	500	4	US-09-265-630-13
c 32	170.5	0.3	52	4	US-09-227-357-537
c 33	168	0.3	388	4	US-09-265-630-11
34	166	0.3	72	4	US-09-227-357-655
c 35	165.5	0.3	3200	2	US-08-477-451-8
c 36	160	0.3	3287	2	US-08-477-451-7
37	156.5	0.3	1388	4	US-09-572-191-2
38	156.5	0.3	1388	4	US-09-723-262-2
39	156.5	0.3	1388	4	US-09-723-219-2
40	156.5	0.3	1886	4	US-08-938-105-3
c 41	146.5	0.3	944	4	US-09-134-001C-4352
42	146.5	0.3	976	4	US-09-104-324B-4
43	144.5	0.3	1104	4	US-08-923-992A-4
44	144.5	0.3	1354	3	US-08-685-871-2
45	145	0.3	1388	2	US-08-685-576-1

ALIGNMENTS

RESULT 1

US-09-457-066-37 Application US/09457066
Sequence 37, Appli
Patent No. 32673

GENERAL INFORMATION:

APPLICANT: Gao, Zeren
APPLICANT: Hart, Charles E.
APPLICANT: Piddington, Christopher S.
APPLICANT: Sheppard, Paul O.
APPLICANT: Shoemaker, Kimberly E.
APPLICANT: Gilbertson, Debra G.
APPLICANT: West, James W.
TITLE OF INVENTION: GROWTH FACTOR, HOMOLOG ZVEGF3
FILE REFERENCE: 98/601,441
CURRENT APPLICATION NUMBER: US/09/457-066
CURRENT FILING DATE: 1999-12-07
NUMBER OF SEQ ID NOS: 50
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 37

LENGTH: 370

TYPE: PRT

ORGANISM: Homo sapiens

US-09-457-066-37

Alignment Scores:

Pred. No.: 6.98e-32

Score: 409-00

Percent Similarity: 100-00%

Best Local Similarity: 100-00%

Query Match: 0-78%

DB: 4

Length: 370

Matches: 73

Conservative: 0

Mismatches: 0

Indels: 0

Gaps: 0

US-10-083-853B-2 (1-29921) x US-09-457-066-37 (1-370)

Qy

17952 GTTGACCTGGATAGGCTCAATGATGCCAAGCGTTACAGTTGCTCCAGCAATTAC 17893
|||||

Db 258 ValAspLeuAspArgLeuAsnAspAlaLysArgTyrSerCysThrProArgAsnTyr 277
QY 17892 TCGGTCAATATAGAGAGAGCTGAAGTTGGCAATGTGCTCTCTTTCCAGTTGCCCTC 17833
Db 278 SerValAsnIleArgGluGluLeuLysLeuAlaAsnValAlaPhePheProArgCysLeu 297
QY 17832 CTCGTCCAGCGCTGTGGAGAAATGTGCTGTGGAACGTGCAACTGGAGGTCCTGCACA 17773
Db 298 LeuValGlnArgCysGlyAsnCysGlyCysGlyThrValAsnTrpArgSerCysThr 317
QY 17772 TGCATTCAGGGAACCGTGAAGAGTATCATGAGGTA 17734
Db 318 CysAsnSerGlyLysThrValLysLysTyrHisGluVal 330

RESULT 2

US-09-540-224-2

; Sequence 22, Application US/09540224

; Patent No. 6468543

; GENERAL INFORMATION:

; APPLICANT: Gilbertson, Debra G.

; APPLICANT: Hart, Charles E.

; TITLE OF INVENTION: METHODS FOR PROMOTING GROWTH OF BONE,

; FILE REFERENCE: 00-28

; CURRENT APPLICATION NUMBER: US/09/540,224

; CURRENT FILING DATE: 2000-03-31

; EARLIER APPLICATION NUMBER: US 60/180,169

; EARLIER FILING DATE: 2000-02-04

; NUMBER OF SEQ ID NOS: 9

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 2

; LENGTH: 370

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-540-224-2

Alignment Scores:

Pred. No.:	6,98e-32	Length:	370
Score:	409.00	Matches:	73
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	4	Gaps:	0

US-10-083-853b-2 (1-29921) x US-09-540-224-2 (1-370)

QY 17952 GTTGACCTGGATAGCTCAATGATGATGCAAGCGTTACAGTTGCACCTCCAGGAATTAC 17893
Db 258 ValAspLeuAspArgLeuAsnAspAlaLysArgTyrSerCysThrProArgAsnTyr 277
QY 17892 TCGGTCAATATAGAGAGAGCTGAAGTTGGCAATGTGCTCTCTTTCCAGTTGCCCTC 17833
Db 278 SerValAsnIleArgGluGluLeuLysLeuAlaAsnValAlaPhePheProArgCysLeu 297
QY 17832 CTCGTCCAGCGCTGTGGAGAAATGTGCTGTGGAACGTGCAACTGGAGGTCCTGCACA 17773
Db 298 LeuValGlnArgCysGlyAsnCysGlyCysGlyThrValAsnTrpArgSerCysThr 317
QY 17772 TGCATTCAGGGAACCGTGAAGAGTATCATGAGGTA 17734
Db 318 CysAsnSerGlyLysThrValLysLysTyrHisGluVal 330

RESULT 3

US-09-540-224-4

; Sequence 4, Application US/09540224

; Patent No. 6468543

; GENERAL INFORMATION:

; APPLICANT: Gilbertson, Debra G.

; APPLICANT: Hart, Charles E.

; TITLE OF INVENTION: METHODS FOR PROMOTING GROWTH OF BONE,

; FILE REFERENCE: 00-28

; CURRENT APPLICATION NUMBER: US/09/540,224

; CURRENT FILING DATE: 2000-03-31
; EARLIER APPLICATION NUMBER: US 60/180,169
; EARLIER FILING DATE: 2000-02-04
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 370
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-540-224-4

Alignment Scores:

Pred. No.:	3,79e-29	Length:	370
Score:	382.00	Matches:	66
Percent Similarity:	95.89%	Conservative:	4
Best Local Similarity:	90.41%	Mismatches:	3
Query Match:	0.73%	Indels:	0
DB:	4	Gaps:	0

US-10-083-853b-2 (1-29921) x US-09-540-224-4 (1-370)

QY 17952 GTTGACCTGGATAGCTCAATGATGATGCAAGCGTTACAGTTGCACCTCCAGGAATTAC 17893
Db 258 ValAspLeuAspArgLeuAsnAspAlaLysArgTyrSerCysThrProArgAsnTyr 277
QY 17892 TCGGTCAATATAGAGAGAGCTGAAGTTGGCAATGTGCTCTCTTTCCAGTTGCCCTC 17833
Db 278 SerValAsnIleArgGluGluLeuLysLeuAlaAsnValAlaPhePheProArgCysLeu 297
QY 17832 CTCGTCCAGCGCTGTGGAGAAATGTGCTGTGGAACGTGCAACTGGAGGTCCTGCACA 17773
Db 298 LeuValGlnArgCysGlyAsnCysGlyCysGlyThrValAsnTrpArgSerCysThr 317
QY 17772 TGCATTCAGGGAACCGTGAAGAGTATCATGAGGTA 17734
Db 318 CysSerSerGlyLysThrValLysLysTyrHisGluVal 330

RESULT 4

US-09-058-489-22

; Sequence 22, Application US/09058489

; Patent No. 6103886

; GENERAL INFORMATION:

; APPLICANT: Whitehead Institute for Biomedical Research

; APPLICANT: Lahn, Bruce

; APPLICANT: Page, David

; TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of

; FILE REFERENCE: WH197-08pa

; CURRENT APPLICATION NUMBER: US/09/058,489

; CURRENT FILING DATE: 1998-04-10

; EARLIER APPLICATION NUMBER: 60/041,877

; EARLIER FILING DATE: 1997-04-11

; NUMBER OF SEQ ID NOS: 91

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 22

; LENGTH: 1079

; TYPE: PRT

; ORGANISM: Human

US-09-058-489-22

Alignment Scores:

Pred. No.:	6,84e-22	Length:	1079
Score:	314.00	Matches:	63
Percent Similarity:	76.92%	Conservative:	7
Best Local Similarity:	69.23%	Mismatches:	21
Query Match:	0.60%	Indels:	0
DB:	3	Gaps:	0

US-10-083-853b-2 (1-29921) x US-09-058-489-22 (1-1079)

QY 10358 AGTTTCAGTCTCTTGTGCCAGCGTAGTGAATCGTGTGATCCCGGGTCACTGCAACCT 10417
Db 989 SerPheGlnGluSerLeuArgAlaGlyMetGlnTrpCysAspLeuSerSerLeuGlnPro 1008

QY 10418 CCACCTCCGGGTTCAAGTATTCTCTCGCTCAGCTCCCAAGTAGTGGGACTACAGG 10477
||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1009 ProProGlyPheLysArgPheSerHisLeuSerLeuProAsnSerTrpAsnTyrArg 1028
QY 10478 CAGGTGCCACCGCTGCTCTATTTTCTATTATTTAGTAGAGATGGGTTTCCCGACGTT 10537
||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1029 HisLeuProSerCysProThrAsnPheCysIlePheValGluThrGlyPheHisVal 1048
QY 10538 GCGCAGGCTGGTCTTGAACCTCTGACCTCAGCTGATCTCGCTCAGCTCCCAAAAT 10597
||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1049 GlyGlnAlaCysLeuGlnLeuThrSerGlyGlyLeuLeuAlaSerAlaSerGlnSer 1068
QY 10598 GCTGGGATTACAGGATGACCGACTGCGGCCGG 10630
Db 1069 AlaGlyIleThrGlyValSerHisAlaArg 1079

RESULT 5

US-09-605-785-573
; Sequence 573, Application US/09605785
; Patent No. 6321716

GENERAL INFORMATION:
; APPLICANT: Xu, Jianshun

; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.

; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi

; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.

; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.

; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.

; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick

; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun

; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER

; CURRENT APPLICATION NUMBER: US/09/605,785
; CURRENT FILING DATE: 2000-06-27

; NUMBER OF SEQ ID NOS: 835
; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 573
; LENGTH: 132

; TYPE: PRT
; ORGANISM: Homo sapiens

US-09-605-785-573

Alignment Scores:
Pred. No.: 2,62e-22 Length: 132

Score: 311.00 Matches: 66

Percent Similarity: 68.27% Conservative: 5

Best Local Similarity: 63.46% Mismatches: 33

Query Match: 0.59% Indels: 1

DB: 4 Gaps: 0

US-10-083-853B-2 (1-29921) x US-09-605-785-573 (1-132)

QY 10340 TTTTCTTTTTCAGATGGAGTTTCAGTCTCTTGGCCAGGCTAGAGTGAATGGTGTGAT 10399
||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 28 PhePheLeuArgGlnClusSerGlyProValAlaGlnAlaGlyValGlnTrpHisasp 47

QY 10400 CCCGGGTCACTGCAACCTCCACCTCCCGGGTTCAAGTGAATCTCTCGCTCAGCCTCCCA 10459
||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 48 LeuSerSerLeuGlnProLeuProHisArgPheLysGlnPheSerCysLeuSerLeuPro 67

QY 10460 AGTACTGGGACTACAGGACCTGGCCACCACCGCTGGCTGAATTTTGTATTTTAGTAGA 10519
||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 68 HisSerTrpAspHisArgTyrAlaProHisLeuAlaAsnPheCysSerPheSerArg 87

QY 10520 GATGGTTT-TCCCCACGTTTGGCAGGCTGGTCTTGAACCTCTGACCTCAGGTGATCTGCC 10578
||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 88 AspGlyValSerLeuCysSerGlyTrpSerLysThrProGlyLeuGlnGlnSerAla 107

QY 10579 TGCTCAGCTCCCAAAATGCTGGGATTACAGGACATGAGCCACTGCGCCCGGTCTCTTCC 10638
||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 108 CysLeuGlyLeuProLysCysTrpGlyTyrArgHisLysProHisProAlaCysHis 127

QY 10639 TAACTTCTAAAC 10650
||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 128 IleLeuLeuAsn 131

RESULT 6
US-08-454-557C-121

; Sequence 121, Application US/08454557C
; Patent No. 5830670

GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne

; APPLICANT: Wands, Jack R.
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection

; TITLE OF INVENTION: of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.

; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington

; STATE: D.C.
; COUNTRY: U.S.A.

; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/454,557C

; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.

; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609,3840003

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600

; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 121:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 375 amino acids

; TYPE: amino acid
; TOPOLOGY: linear

; MOLECULE TYPE: protein
US-08-454-557C-121

Alignment Scores:
Pred. No.: 1,92e-21 Length: 375

Score: 306.00 Matches: 76

Percent Similarity: 71.07% Conservative: 10

Best Local Similarity: 62.81% Mismatches: 30

Query Match: 0.58% Indels: 5

DB: 2 Gaps: 2

US-10-083-853B-2 (1-29921) x US-08-454-557C-121 (1-375)

QY 10328 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 10387
||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 195 PheIlePheIlePheAsnPheLeuArgGlnSerLeuAsnSerValThrGlnAlaGlyVal 214

QY 10388 CAATGGTGATCCCGGGTCACTGCAACCTCCACCTCCCGGGTTCAGGTGATTCCTCTGC 10447
||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 215 GlnTrpArgAsnLeuGlySerLeuGlnProLeuProGlyPheSerLeuPheSerCys 234

QY 10448 CTCAGCCTCCCAAGTAGCTGGGACTACAGGCACGTGCCACCGCTGGCTAATTTT--- 10505

Db 215 GlnTrpArgAsnLeuGlySerLeuGlnProLeuProProGlyPheLysLeuPheSerCys 234
QY 10448 CTCAGCCTCCACAGTAGCTGGGACTACAGGACGCTGCACACCGCTGCTGCTATTTT-- 10505
Db 235 ProSerLeuLeuSerSerTrpAspTyrArgArg---ProProArgLeuAlaAsn-PhePh 253
QY 10506 -GTATTTTATAGATAGATGGTTTCCACAGCTTGGCCAGGCTGCTTGAACCTCCGTGACC 10564
Db 253 eValPheLeuValGluMetGlyPheThrMetPheAlaArgLeuLeuLeuSerGlyPr 273
QY 10565 TCAGGTGATCTGCTGCTGCTAGCCTCCCAAAATGCTGGGATTACAGGCATGAGCCACTGC 10624
Db 273 o-CysAspLeuProAlaSerAlaSerGlnSerAlaGlyIleThrGlyValSerHisHis 292
QY 10625 GCCCGGCTCTCTCCCTAACTTCTAAACACGCTGCTTTTCAGCACACACACTTCTCAA 10681
Db 293 AlaArgLeuIlePheAsnPheCysLeuPheGluMetGluSerHisValThrGln 311

RESULT 9

PCT-US95-17111A-121
; Sequence 121, Application PC/TUS9517111A
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and
; TITLE OF INVENTION: Detection of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/17111A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/340,426
; FILING DATE: 14-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 121:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 397 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-17111A-121

Alignment scores:
Pred. No.: 8 64e-19 Length: 397
Score: 280.00 Matches: 65
Percent Similarity: 76.29% Conservative: 9
Best Local Similarity: 67.01% Mismatches: 21
Query Match: 0.54% Indels: 3
DB: 5 Gaps: 0

US-10-083-853B-2 (1-29221) x PCT-US95-17111A-121 (1-397)

QY 10333 TTTTNTTTTTCAGATGGAGTTTCAGTCTTCTTGCCAGGCTAGAGTGAATG 10392

Db 298 PheAsnPheCysLeuPheGluMetGluSerHisSer-ValThrGlnAlaGlyValGlnTr 317
QY 10393 GTGTGATCCCGGGTCACTGCAACCTCCACCTCCCGGGTTCACAGTGTTCCTCCGCTCAG 10452
Db 317 pProAsnLeuGlySerLeuGlnProLeuProGlyLeuLysArgPheSerCysLeuSe 337
QY 10453 CTTCCCAACTAGCTGGGACTACAGGCACGTCGCCACACCGCTGGCTAAATTTTGTATTTT 10512
Db 337 rLeuProSerSerTrpAspTyrGlyHisLeu-HisHisThrProLeuIlePheValPheS 357
QY 10513 TAGTAGAGATGGTTTCCC-CACGTTGGCCAGGCTGCTTGAACCTCTGACCTCAGGTG 10571
Db 357 erLeuGluAlaGlyPheHisHisIleCysGlnAlaGlyLeuLysLeuLeuThrSerGlyA 377
QY 10572 ATCTGCCTGCCCTCAGCTCCCAAAATGCTGGGATTACAGGCATGACC 10618
Db 377 spProProAlaSerAlaPheGlnSerAlaGlyIleThrGlyValThr 392

RESULT 10

US-09-227-357-193
; Sequence 193, Application US/09227357
; Patent No. 6342581
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: P2010P1
; CURRENT APPLICATION NUMBER: US/09/227,357
; CURRENT FILING DATE: 1999-01-08
; EARLIER APPLICATION NUMBER: PCT/US98/13684
; EARLIER FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: 60/051,926
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,793
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,925
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,929
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,803
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,732
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,931
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,932
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,916
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,930
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,918
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,920
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,733
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,795
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,919
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,928
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/055,722
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,723
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,948
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,949
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,953
; EARLIER FILING DATE: 1997-08-18

EARLIER APPLICATION NUMBER: 60/051,918
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,920
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,733
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,795
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,919
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,928
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/055,722
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,723
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,948
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,949
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,953
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,950
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,947
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,964
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/056,360
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,684
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,984
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,954
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/058,785
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,664
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,660
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,661
EARLIER FILING DATE: 1997-09-12
NUMBER OF SEQ ID NOS: 672
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 285
TYPE: PRT
ORGANISM: Homo sapiens
US-09-227-357-285

Alignment Scores:
Pred. No.: 1,85e-16 Length: 122
Score: 253.00 Matches: 51
Percent Similarity: 74.32% Conservative: 4
Best Local Similarity: 68.92% Mismatches: 19
Query Match: 0.48% Indels: 0
DB: 4 Gaps: 0

US-10-083-853B-2 (1-29921) x US-09-227-357-285 (1-122)

QY 10409 CTGCAACCTCCACCTCCCGGGTCAAGTCTCTGCTGCTCCAGCTCCCAAGTAGCTGG 10468
Db 1 MetGlnAlaLeuProProGlyPheLysGlnPheSerCysLeuSerLeuProSerArgTtp 20
QY 10469 GACTACAGGACGTGCCACCCAGCCCTGGCTAATTTTGTATTTTAGTAGAGATGGGTTT 10528
Db 21 AspTyrGlyCysAlaThrGlnHisProAlaAsnPheCysIlePheArgAspArgVal 40
QY 10529 CCCACGTGGCCAGCGCTGGTGTGAACCTCCAGCTCAGGTGATCTGCTCGCTCAGCC 10589
Db 41 SerHisValGlyGlnAlaGlyLeuLysLeuLeuThrSerValAspProProAlaTtpAla 60

QY 10589 TCCCAAAATCGTGGGATTACAGGCATGAGCCACTGGCGCCGG 10630
Db 61 SerGlnSerAlaGlyIleThrGlyLysSerHisCysAlaGln 74

RESULT 13

US-09-040-220D-2
Sequence 2, Application US/09040220D
Patent No. 6391311
GENERAL INFORMATION:
APPLICANT: Ferrara, Napoleone
APPLICANT: Kuo, Sophia S.
TITLE OF INVENTION: NOVEL POLYPEPTIDES HAVING HOMOLOGY TO VASCULAR
TITLE OF INVENTION: ENDOTHELIAL CELL GROWTH FACTOR AND BONE MORPHOGENETIC
TITLE OF INVENTION: PROTEIN 1 AND NUCLEIC ACIDS ENCODING SAME, THEIR USES,
TITLE OF INVENTION: AND PROCESSES FOR THEIR PRODUCTION
FILE REFERENCE: P1122
CURRENT APPLICATION NUMBER: US/09/040,220D
CURRENT FILING DATE: 1998-03-17
NUMBER OF SEQ ID NOS: 8
SEQ ID NO 2
LENGTH: 345
TYPE: PRT
ORGANISM: Human
US-09-040-220D-2

Alignment Scores:
Pred. No.: 1.1e-14 Length: 345
Score: 239.00 Matches: 42
Percent Similarity: 71.23% Conservative: 10
Best Local Similarity: 57.53% Mismatches: 21
Query Match: 0.46% Indels: 0
DB: 4 Gaps: 0

US-10-083-853B-2 (1-29921) x US-09-040-220D-2 (1-345)

QY 17952 GTTGACCTGGATAGGCTCAATGATGATGCAAGCGTTACAGTTGCTACTCCAGGAATTAC 17893
Db 236 ValAspLeuAsnLeuLeuThrGluGluValArgLeuTyrSerCysThrProArgAsnPhe 255
QY 17892 TCGGTCAATATAGAGAAGAGCTGAAGTTGGCCAAATGTGGTCTCTTCCACGTTCCCTC 17833
Db 256 SerValSerIleArgGluLeuLysArgThrAspThrIlePheThrProGlyCysLeu 275
QY 17832 CTGCTGACGCGCTGTGAGGAATCTGCTGTGGAACCTCACTGAGGCTCCTGCACA 17773
Db 276 LeuValLysArgCysGlyGlyAsnCysAlaCysCysLeuHisAsnCysAsnGluCysGln 295
QY 17772 TGCATTCAGGGAACCCGTGAAAAAGTATCATGAGGTA 17734
Db 296 CysValProSerLysValThrLysLysTyrHisGluVal 308

RESULT 14

US-09-457-066-2
Sequence 2, Application US/09457066
Patent No. 6432673
GENERAL INFORMATION:
APPLICANT: Gao, Zeren
APPLICANT: Hart, Charles E.
APPLICANT: Piddington, Christopher S.
APPLICANT: Sheppard, Paul O.
APPLICANT: Shoemaker, Kimberly E.
APPLICANT: Gilbertson, Debra G.
APPLICANT: West, James W.
TITLE OF INVENTION: GROWTH FACTOR HOMOLOG ZVEGF3
FILE REFERENCE: 98-60
CURRENT APPLICATION NUMBER: US/09/457,066
CURRENT FILING DATE: 1999-12-07
NUMBER OF SEQ ID NOS: 50
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 345
TYPE: PRT

; ORGANISM: Homo sapiens
US-09-457-066-2

Alignment Scores:

Pred. No.:	1.1e-14	Length:	345
Score:	239.00	Matches:	42
Percent Similarity:	71.23%	Conservative:	10
Best Local Similarity:	57.53%	Mismatches:	21
Query Match:	0.46%	Indels:	0
DB:	4	Gaps:	0

US-10-083-853B-2 (1-29921) x US-09-457-066-2 (1-345)

Qy	17952	GTTCACCTGGATAGCTCAATGATGCCAAGCGTTACAGTTGCACCTCCAGGAATTAC	17893
Db	236	ValAspLeuAsnLeuLeuThrGluGluValArgLeuTyrSerCysThrProArgAsnPhe	255
Qy	17892	TCGGTCAATATAAGAGAGAGACTGAAGTTGGCCAAATGGTCTTCTTCCACAGTTGCCTC	17833
Db	256	SerValSerIleArgGluGluLeuLysArgThrIlePheThrIleProGlyCysLeu	275
Qy	17832	CTCGTCAGCGCTGTGGAGGAAATGTGGCTGTGGAACGTGCAACTGGAGTCTCTGCACA	17773
Db	276	LeuValLysArgCysGlyGlyAsnCysAlaCysCysLeuHisAsnCysAsnGluCysGln	295
Qy	17772	TGCAATTCAGGAAACCGTGAAGAAAGTATCATGAGGTA	17734
Db	296	CysValProSerLysValThrLysLysTyrHisGluVal	308

RESULT 15

US-09-265-686-2

; Sequence 2, Application US/09265686
; Patent No. 6455283
; GENERAL INFORMATION:
; APPLICANT: Ferrara, Napoleone
; TITLE OF INVENTION: POLYPEPTIDES HOMOLOGOUS TO VEGF AND BMP1
; FILE REFERENCE: P1122P2
; CURRENT APPLICATION NUMBER: US/09/265,686
; PRIOR FILING DATE: 1999-03-10
; PRIOR APPLICATION NUMBER: US 09/040,220
; PRIOR FILING DATE: 1998-03-17
; PRIOR APPLICATION NUMBER: US 09/184,216
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 2
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Human
US-09-265-686-2

Alignment Scores:

Pred. No.:	1.1e-14	Length:	345
Score:	239.00	Matches:	42
Percent Similarity:	71.23%	Conservative:	10
Best Local Similarity:	57.53%	Mismatches:	21
Query Match:	0.46%	Indels:	0
DB:	4	Gaps:	0

US-10-083-853B-2 (1-29921) x US-09-265-686-2 (1-345)

Qy	17952	GTTCACCTGGATAGCTCAATGATGCCAAGCGTTACAGTTGCACCTCCAGGAATTAC	17893
Db	236	ValAspLeuAsnLeuLeuThrGluGluValArgLeuTyrSerCysThrProArgAsnPhe	255
Qy	17892	TCGGTCAATATAAGAGAGAGACTGAAGTTGGCCAAATGGTCTTCTTCCACAGTTGCCTC	17833
Db	256	SerValSerIleArgGluGluLeuLysArgThrIlePheThrIleProGlyCysLeu	275
Qy	17832	CTCGTCAGCGCTGTGGAGGAAATGTGGCTGTGGAACGTGCAACTGGAGTCTCTGCACA	17773
Db	276	LeuValLysArgCysGlyGlyAsnCysAlaCysCysLeuHisAsnCysAsnGluCysGln	295

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 15:09:32 ; Search time 1095.73 Seconds

(without alignments)

7405.060 Million cell updates/sec

Title: US-10-083-853B-2_COPY_29421_29921

Perfect score: 501

Sequence: I ctttgggttttagacatg.....caattctgtgaagaagtaa 501

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: EST

1: em_estba:*

2: em_estin:*

3: em_estm:*

4: em_estov:*

5: em_estpl:*

6: em_estro:*

7: em_estl:*

8: em_estl:*

9: gb_estl:*

10: gb_est2:*

11: gb_est3:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: gb_gss:*

18: em_gss_hum:*

19: em_gss_inv:*

20: em_gss_pln:*

21: em_gss_vrt:*

22: em_gss_fun:*

23: em_gss_mam:*

24: em_gss_mus:*

25: em_gss_other:*

26: em_gss_pro:*

27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	470.6	93.9	849	9	AU118935	AU118935 AU118935
C 2	466.4	93.1	658	13	BM509719	BM509719 ig93b09.y
C 3	465.8	93.0	750	9	AU119706	AU119706 AU119706
C 4	465.8	93.0	1052	11	BC032807	BC032807 Homo sapi
C 5	465.8	93.0	2776	11	BC028293	BC028293 Homo sapi
C 6	464.2	92.7	760	13	BI092874	BI092874 602857901

7	464.2	92.7	1069	13	BM549258	BM549258 AGENCOURT
8	462.6	92.3	673	17	AG102893	AG102893 Pan trogl
C 9	461	92.0	681	10	AW850541	AW850541 IL3-CT021
C 10	459.4	91.7	602	9	AL711636	AL711636 DKF2p686N
C 11	459.4	91.7	695	17	AQ346406	AQ346406 RPCI11-11
C 12	458.4	91.5	1049	14	BQ071465	BQ071465 AGENCOURT
C 13	458.2	91.5	804	12	BG288152	BG288152 602387957
C 14	457.8	91.4	679	17	AQ491435	AQ491435 RPCI-11-2
C 15	457.2	91.3	656	17	AG077683	AG077683 Pan trogl
C 16	457.0	90.8	579	10	AW845700	AW845700 MRO-CT006
C 17	454.6	90.7	805	12	BG680574	BG680574 602629008
C 18	453.8	90.6	1040	12	BE869508	BE869508 601446175
C 19	453	90.4	650	17	AG047226	AG047226 Pan trogl
C 20	453	90.4	840	12	BG623383	BG623383 602648229
C 21	453	90.4	954	12	AQ748961	AQ748961 HS_5574_A
C 22	452.2	90.3	954	12	BG104752	BG104752 602311892
C 23	452	90.2	717	17	AQ052948	AQ052948 RPCI11-49
C 24	451.2	90.1	2703	17	AF101616	AF101616 AF101616
C 25	450.6	89.9	778	13	BI084663	BI084663 602869851
C 26	449	89.6	876	14	BQ932611	BQ932611 AGENCOURT
C 27	448.4	89.5	736	12	BE788398	BE788398 601480244
C 28	447.8	89.4	789	17	AQ740860	AQ740860 HS_2274_A
C 29	446.8	89.2	522	17	AQ354733	AQ354733 CTBTI-EL
C 30	446.8	89.2	659	17	B83611	B83611 RPCI11-16F1
C 31	446.4	89.1	789	13	BI084336	BI084336 602869851
C 32	445.8	89.0	915	12	BG171891	BG171891 602320924
C 33	445.8	89.0	989	12	BG679572	BG679572 602627615
C 34	445.4	88.9	691	17	AG087253	AG087253 Pan trogl
C 35	445.4	88.9	747	12	BG548021	BG548021 602575934
C 36	445.2	88.9	1821	17	AF101873	AF101873 AF101873
C 37	444.8	88.8	863	13	BI257428	BI257428 602967772
C 38	444	88.6	690	9	AL707561	AL707561 DKF2p6861
C 39	443.2	88.5	908	12	BG353707	BG353707 602563823
C 40	442.4	88.3	820	12	BG621601	BG621601 602617011
C 41	442	88.2	517	17	B64761	B64761 CIT-HSP-200
C 42	441.8	88.2	744	17	AQ393258	AQ393258 CTBTI-EL
C 43	441	88.0	999	14	BM905337	BM905337 AGENCOURT
C 44	440.6	87.9	679	14	BM972622	BM972622 UI-CF-EC1
C 45	440.4	87.9	890	17	AQ739555	AQ739555 HS_5381_B

ALIGNMENTS

RESULT 1	AU118935/C	AU118935	849 bp	mRNA	linear	EST 01-AUG-2002
LOCUS	AU118935	HEMBAL Homo sapiens	cdNA clone	HEMBAL004677	5', mRNA	
DEFINITION	AU118935	sequence.				
ACCESSION	AU118935					
VERSION	AU118935.1	GI:10934170				
KEYWORDS	EST.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
AUTHORS	Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y., Yamamoto,C., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and Isogai,T.					
TITLE	HRI human cDNA project					
JOURNAL	Unpublished (2000)					
COMMENT	Genomics Laboratory Helix Research Institute 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan Contact: Takao Isogai Tel: 81-438-52-3986 Fax: 81-438-52-3986 Email: genomics@hri.co.jp HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.					

```
FEATURES             Location/Qualifiers
  source              1..849
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="HEMBA1004677"
                    /clone_lib="HEMBA1"
                    /tissue_type="whole embryo, mainly head"
                    /dev_stage="embryo, 10 weeks"
                    /note="Vector: pME18FL3"
BASE COUNT          351 a 190 c 137 g 167 t      4 others
ORIGIN
Query Match          93.98; Score 470.6; DB 9; Length 849;
Best Local Similarity 96.28; Pred. No. 7.5e-106;
Matches 482; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
Qy 1 CTTTGTGTTTGTAGACATGAAGTCCTTGCCCATGCTATGCTGTAATGGTATGCGCTG 60
    |||||
Db 525 CTTTGTGTTTGGACATGAAGTCCTTGCCCATGCTATGCTGTAATGGTATGCGCTA 466
Qy 61 GGTTCCTCTAGGGATTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTT 120
    |||||
Db 465 GGTTCCTCTAGGGATTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTT 406
Qy 121 AATTAATTTTGTATAGGTGTAGGAAGGATCCAGTTTCAGCTTTTCTACATAGGGCTA 180
    |||||
Db 405 AATTTGATTTTGTATAGGTGTAGGAAGGATCCAGTTTCAGCTTTTCTACATAGGGCTA 346
Qy 181 GCCAGTTTCTCAGCACCAATTTAATAATAGGGAATCCTTCCCATGCTCTGTTTCT 240
    |||||
Db 345 GCCAGTTTCTCAGCACCAATTTAATAATAGGGAATCCTTCCCATGCTCTGTTTCT 286
Qy 241 CAGGTTTGTCAAAGATCAGATAGTTGTAGATATGCGGCATTTATCTCGAGGGCTCTGTT 300
    |||||
Db 285 CAGGTTTGTCAAAGATCAGATAGTTGTAGATATGCGGCATTTATCTCGAGGGCTCTGTT 226
Qy 301 TGTTCATTTGGTGAATCTCTGTTTGTGGTACCAGTACCAGTGTGTTTGGTACTGTAG 360
    |||||
Db 225 TGTTCATTTGATCTATCTCTGTTTGTGGTACCAGTACCAGTGTGTTTGGTACTGTAG 166
Qy 361 CTTGTAGTGTAGTTTCAAGTCAGGTAGATGATGCTCCAGCTTTGCTCTTTGGCTTA 420
    |||||
Db 165 CTTGTAGTGTAGTTTCAAGTCAGGTAGTGTGATGCTCCAGCTTTGCTCTTTGGCTTA 106
Qy 421 GGATGACTTGGCGATCGGCTCTTTTGTGGTCCACATGAACCTTTAAAGTAGTTTTT 480
    |||||
Db 105 GGATGACTTGGCAATCGCGCTCTTTTGTGGTCCATATGAACCTTTAAAGTAGTTTTT 46
Qy 481 CCAATCTGTGAAGAAGTAA 501
    |||||
Db 45 CCAATCTGTGAAGAAGTCA 25

RESULT 2
BM509719/c          658 bp      mRNA      linear      EST 15-FEB-2002
LOCUS              i93309.y1 HR85 islet Homo sapiens cDNA 5' similar to TR:Q14288
DEFINITION          Q14288 HYPOTHETICAL PROTEIN ;, mRNA sequence.
ACCESSION           BM509719
VERSION             BM509719.1 GI:18680862
KEYWORDS             EST.
SOURCE              human.
ORGANISM            Homo sapiens
                    Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
                    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE           1 (bases 1 to 658)
AUTHORS             Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
                    Lemishka,I., Seearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
                    Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
                    Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
                    M., Gibbons,M., McCann,R., Cole,R., Tsagarelshvili,R., Williams,T.,
                    Jackson,Y. and Bowers,Y.
                    Endocrine Pancreas Consortium
TITLE               CCAATCTGTGAAGAAGTCA 87
```

```
JOURNAL             Unpublished (2000)
COMMENT             Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
                    Endocrine Pancreas Consortium
                    Harvard University, Howard Hughes Medical Institute
                    Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
                    MA 02138
                    Tel: 617-495-1812
                    Fax: 617-495-8557
                    Email: dmelton@biohph.harvard.edu
                    Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
                    Washington University Genome Sequencing Center For information on
                    obtaining a clone please contact: Dr. Hiroshi Inoue
                    (hinoue@im.wustl.edu)
                    Seq primer: -40RP from Gibco
                    High quality sequence stop: 481.
FEATURES             Location/Qualifiers
  source              1..658
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone_lib="HR85 islet"
                    /tissue_type="Purified pancreatic islet"
                    /lab_host="DH10B"
                    /note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1:
                    NotI; Site_2: XhoI; cDNA made by oligo-dT priming.
                    Size-selected on agarose gel. Average insert size ~1kb. 5'
                    XhoI site was destroyed after directional cloning.
                    Amplified once. Contact information: Hiroshi Inoue, MD,
                    Metabolism Div. (Alan Permutt Lab), Washington University
                    School of Medicine, Box 8127, 660 South Euclid Ave., St.
                    Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:
                    314-362-1916, Fax: 314-747-2692."
BASE COUNT          261 a 142 c 120 g 134 t      1 others
ORIGIN
Query Match          93.18; Score 466.4; DB 13; Length 658;
Best Local Similarity 95.68; Pred. No. 7.6e-105;
Matches 479; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
Qy 1 CTTTGTGTTTGTAGACATGAAGTCCTTGCCCATGCTATGCTGTAATGGTATGCGCTG 60
    |||||
Db 587 CTTTGTGTTTGTAGACATGAAGTCCTTGCCCATGCTATGCTGTAATGGTATGCGCTA 528
Qy 61 GGTTCCTCTAGGGATTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTT 120
    |||||
Db 527 GGTTCCTCTAGGGATTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTT 468
Qy 121 AATTAATTTTGTATAGGTGTAGGAAGGATCCAGTTTTCAGCTTTCTACATAGGGCTA 180
    |||||
Db 467 AATTAATTTTGTATAGGTGTAGGAAGGATCCAGTTTTCAGCTTTCTACATAGGGCTA 408
Qy 181 GCCAGTTTCTCAGCACCAATTTAATAATAGGGAATCCTTTCGCCATGCTGTTTCT 240
    |||||
Db 407 GCCAGTTTCTCAGCACCAATTTAATAATAGGGAATCCTTTCGCCATGCTGTTTCT 348
Qy 241 CAGGTTTGTCAAAGATCAGATAGTTGTAGATATGCGGCATTTATCTCGAGGGCTCTGTT 300
    |||||
Db 347 CAGGTTTGTCAAAGATCAGATAGTTGTAGATATGCGGCATTTATCTCGAGGGCTCTGTT 288
Qy 301 TGTTCATTTGGTGTATATCTCTGTTTGTGGTACCAGTACCAGTGTGTTTGTGGTACTGTAG 360
    |||||
Db 287 TGTTCATTTGATCTATATCTCTGTTTGTGGTACCAGTACCAGTGTGTTTGTGGTACTGTAG 228
Qy 361 CTTGTAGTGTAGTTTGAAGTCAGGTAGCATGATGCTCCAGCTTTGTTGTTCTTTTGGCTTA 420
    |||||
Db 227 CTTGTAGTGTAGTTTGAAGTCAGGTAGCATGATGCTCCAGCTTTGTTGTTCTTTTGGCTTA 168
Qy 421 GGATGACTTGGCGATGTGGCTCTTTTGTGGTCCACATGAACCTTTAAAGTAGTTTTT 480
    |||||
Db 167 GGATGACTTGGTGTATGCGGGCTCTTTTGTGGTCCATATGAACCTTTAAAGTAGTTTTT 108
Qy 481 CCAATCTGTGAAGAAGTAA 501
    |||||
Db 107 CCAATCTGTGAAGAAGTCA 87
```

```
RESULT 3
AUI19706/c
LOCUS
DEFINITION
AUI19706 HEMBA1 Homo sapiens cDNA clone HEMBA1006448 5', mRNA
sequence:
AUI19706
VERSION
AUI19706.1 GI:10934941
EST.
KEYWORDS
human.
SOURCE
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 750)
Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
HRI human cDNA project
Unpublished (2000)
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1332-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project: 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
Location/Qualifiers
1..750
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEMBA1006448"
/clone_lib="HEMBA1"
/tissue_type="whole embryo, mainly head"
/dev_stage="embryo, 10 weeks"
/notes="vector: pME18SFL3"
BASE COUNT 302 a 166 c 123 g 156 t 3 others
ORIGIN
Query Match 93.0%; Score 465.8; DB 9; Length 750;
Best Local Similarity 95.6%; Pred. No. 1.1e-104;
Matches 479; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
Qy 1 CTTTGTGTTTGTAGACATGAAGTCCTTGCCATGCGCTATGCTCTGAATGGTATGCGCTG 60
Db 574 CTTTGTGTTTGTGACATGAAGTCCTTGCCAGCGCTATGCTCTGAATGGTATGCGCTA 515
Qy 61 GGTTCCTCTAGGCAATTTATGTTTATGCTCTAACATTTAAGTCTTTAATCCATCTTG 120
Db 514 GGTTCCTCTAGGCTTTTATGTTTATGTTTATGTTTAACTTTAATCTTTAATCCATCTTG 455
Qy 121 AATTAATTTTGTATAGGTGTAAGGAAGGATCCAGTTTCAGCTTTCTACATAGGCGTA 180
Db 454 AATTGATTTTGTATAGGTGTAAGGAAGGATCCAGTTTCAGCTTTCTACATAGGCGTA 395
Qy 181 GCCAGTTTCTCAGCACCATTTATTAATAGGGAATCCTTCCCATTCGTTGTTTCT 240
Db 394 GCCAGTTTCTCAGCACCATTTATTAATAGGGAATCCTTCCCATTCGTTGTTTCT 335
Qy 241 CAGGTTTGTCAAAGATCAGATAGTTGTAGATATCGCGCATTTATTTCTGAGGCGCTCTTC 300
Db 334 CAGGTTTGTCAAAGATCAGATAGTTGTAGATATCGCGCATTTATTTCTGAGGCGCTCTTC 275
Qy 301 TGTTCATGTTGGTATATCTCTGTTTGGTACCAGTACCAGTATGTTGTTTGGTACTAG 360
Db 274 TGTTCATGATCATATATCTCTGTTTGGTACCAGTACCAGTATGTTTGGTACTAG 215
Qy 361 CTTTGTAGTTAGTTTGAAGTCAGGTAGCATGATGCTCCAGCTTGTCTTTTGGCTTA 420
Db 1 CTTTGTGTTTGTAGACATGAAGTCCTTTGCCATGCGCTATGCTCTGAATGGTATGCGCTG 60

Db 214 CTTGTAGTATAGTTTGAAGTCAGTAGTGATGTCCTCCAGCTTGTCTTTTGGCTTA 155
Qy 421 GGATTGACTTGGCGATGTGGGCTCTTTTTCCTCCACATCACTTTAAAGTAGTATTTT 480
Db 154 GGATTGACTTGGCAATGCGGCTCTTTTTCCTCATATGACATTTAAAGTAGTATTTT 95
Qy 481 CCAATTCCTGTGAAGAAAGTAA 501
Db 94 CCAATTCCTGTGAAGAAAGTCA 74

RESULT 4
BC032807/c
LOCUS
DEFINITION
BC032807 Homo sapiens, clone IMAGE:4480714, mRNA.
ACCESSION
BC032807.1 GI:21618542
VERSION
BC032807.1
KEYWORDS
HTC.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1052)
Straussberg,R.
Direct Submission
Submitted (06-JUN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NH-MGC Project-URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: gcgaps-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc_mgc@nhgri.nih.gov
Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.S., Breen,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
Tsurgou,C., Votz,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Young,A., Zhang,L.-H. and Green,E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 68 Row: f Column: 17
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis
This clone has the following problem: frame shifted.
FEATURES
source
1..1052
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4480714"
/tissue_type="Bladder, transitional cell papilloma"
/clone_lib="NIH_MGC_93"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"
BASE COUNT 434 a 209 c 182 g 227 t
ORIGIN
Query Match 93.0%; Score 465.8; DB 11; Length 1052;
Best Local Similarity 95.6%; Pred. No. 1.2e-104;
Matches 479; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
Qy 1 CTTTGTGTTTGTAGACATGAAGTCCTTTGCCATGCGCTATGCTCTGAATGGTATGCGCTG 60
```

```

Db      855 CTTTGTGTTTATAGACATGAAGTCCCTGCCCCTATGCTCTGAATGATGCTTA 796
Qy      61 GGTTCCTCTAGGATTTTATGTTTATAGTCTTAACATTTAAGTCTTTAATCCATCTTG 120
Db      795 GGTTCCTCTAGGATTTTATGATGTTAGTCTTAACATTTAAGTCTTTAATCCATCTTG 736
Qy      121 AATTAATTTTGTATAGGTGTAAGGAAGGATCCAGTTTCAGCTTTTCTACATAGGCTTA 180
Db      735 AATTGATTTTGTATAGGTGTAAGGAAGGATCCAGTTTCAGCTTTTCTACATAGGCTTA 676
Qy      181 GCCAGTTTCTCAGCACCATTATTAATAAGGAATCCCTTCCCATGCTGTTTCTTCT 240
Db      675 GCCAGTTTCTCAGCACCATTATTAATAAGGAATCCCTTCCCATGCTGTTTCTTCT 616
Qy      241 CAGTTTGTCAAGATCAGATAGTTGTAGATATGCGGCATTTTCTGAGGGCTCTGTTTC 300
Db      615 CAGTTTGTCAAGATCAGATAGTTGTAGATATGCGGCATTTTCTGAGGGCTCTGTTTC 556
Qy      301 TGTTCATGTTGATATCTCTGTTTGGTACCAGTACCATTGTTTGGTGGTACTGTAG 360
Db      555 TGTTCATGTTGATATCTCTGTTTGGTACCAGTACCATTGTTTGGTGGTACTGTAG 496
Qy      361 CCTGTAGTCTAGTTGAAGTCAAGTATGATGATGCTCCAGCTTTGTTTGGGCTTA 420
Db      495 CCTGTAGTCTAGTTGAAGTCAAGTATGATGATGCTCCAGCTTTGTTTGGGCTTA 436
Qy      421 GGATTTGCTGGGATGCTGGCTCTTTTGGTTCACATGAACATTTAAAGTATGTTTCT 480
Db      435 GGATTTGCTGGGATGCTGGCTCTTTTGGTTCACATGAACATTTAAAGTATGTTTCT 376
Qy      481 CCAATTCCTGTGAAGAAAGTAA 501
Db      375 CCAATTCCTGTGAAGAAAGTCA 355

```

RESULT 5

```

BC028293/c
LOCUS      Homo sapiens, clone IMAGE:4704802, mRNA.
DEFINITION BC028293
ACCESSION  BC028293
VERSION    BC028293.1 GI:22418059
KEYWORDS   HTC.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 2776)
            Strausberg,R.
            Direct Submission
            Submitted (10-APR-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
            NIH-MGC Project URL: http://mgc.nci.nih.gov
            Contact: MGC help desk
            Email: cgabbs-remail.nih.gov
            Tissue Procurement: DCTD/DTF
            CDNA Library Preparation: Life Technologies, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hqsc.bcm.tmc.edu/cdna/
            Contact: amg@bcm.tmc.edu
            Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
            Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
            Richards, S., Gibbs, R.A.

```

REMARK

```

COMMENT    Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 29 Row: i Column: 19
            This clone has the following problem: incomplete processing.

```

```

FEATURES             Location/Qualifiers
     source            1..2776
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4704802"
                     /tissue_type="Breast, mammary adenocarcinoma."
                     /clone_lib="NIH_MGC_87"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
BASE COUNT          1002 a      618 c      587 g      569 t
ORIGIN
Query Match      93.0%; Score 465.8; DB 11; Length 2776;
Best Local Similarity 95.6%; Pred. No. 1.6e-104;
Matches 479; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
Qy      1 CTTTTGGTGTTTAGACATGAAGTCCCTGCCCCTATGCTCTGAATGATGCTTA 60
Db      1704 CTTTGGTGTTTAGACATGAAGTCCCTGCCCCTATGCTCTGAATGATGCTTA 1645
Qy      61 GGTTCCTCTAGGATTTTATGTTTATAGTCTTAACATTTAAGTCTTTAATCCATCTTG 120
Db      1644 GGTTCCTCTAGGATTTTATGTTTATAGTCTTAACATTTAAGTCTTTAATCCATCTTG 1585
Qy      121 AATTAATTTTGTATAGGTGTAAGGAAGGATCCAGTTTTCAGCTTTTCTACATAGGCTTA 180
Db      1584 AATTAATTTTGTATAGGTGTAAGGAAGGATCCAGTTTTCAGCTTTTCTACATAGGCTTA 1525
Qy      181 GCCAGTTTCTCAGCACCATTATTAATAAGGAATCCCTTCCCATGCTGTTTCTTCT 240
Db      1524 GCCAGTTTCTCAGCACCATTATTAATAAGGAATCCCTTCCCATGCTGTTTCTTCTTGT 1465
Qy      241 CAGTTTGTCAAGATCAGATAGTTGTAGATATGCGGCATTTATTTCTGAGGGCTCTGTTTC 300
Db      1464 CAGTTTGTCAAGATCAGATAGTTGTAGATATGCGGCATTTATTTCTGAGGGCTCTGTTTC 1405
Qy      301 TGTTCATGTTGATATCTCTGTTTGGTACCAGTACCATTGTTGTTGGTGGTACTGTAG 360
Db      1404 TGTTCATGTTGATATCTCTGTTTGGTACCAGTACCATTGTTGTTGGTGGTACTGTAG 1345
Qy      361 CCTGTAGTCTAGTTTGAAGTCAAGTATGATGCTCCAGCTTGTCTTCTTTGGCTTA 420
Db      1344 CCTGTAGTCTAGTTTGAAGTCAAGTATGATGCTCCAGCTTGTCTTCTTTGGCTTA 1285
Qy      421 GGATTTGACTGGCGATGCTGGCTCTTTTGGTTCACATGAACATTTAAAGTATGTTTCT 480
Db      1284 GGATTTGACTGGCGATGCTGGCTCTTTTGGTTCACATGAACATTTAAAGTATGTTTCT 1225
Qy      481 CCAATTCCTGTGAAGAAAGTAA 501
Db      1224 CCAATTCCTGTGAAGAAAGTCA 1204

```

RESULT 6

```

BI092874/c.
LOCUS      BI092874
DEFINITION 602857901F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:4999176 5',
            mRNA sequence.
ACCESSION  BI092874
VERSION    BI092874.1 GI:14511204
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 760)
            NIH-MGC http://mgc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, ph.D.
            Email: cgabbs-remail.nih.gov
            Tissue Procurement: ATCC
            CDNA Library Preparation: Life Technologies, Inc.

```

cDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: L1AM11029 row: a column: 01

High quality sequence stop: 760.

Location/Qualifiers

FEATURES

source

1. 760

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4999176"

/clone_lib="NIH_MGC_10"

/cell_line="MGC36"

/lab_host="DH10B"

/note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;

Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.

Average insert size 1.5 kb. Library prepared by Life

Technologies."

BASE COUNT 313 a 164 c 133 g 150 t

ORIGIN

Query Match 92.7%; Score 464.2; DB 13; Length 760;

Best Local Similarity 95.4%; Pred. No. 2.8e-104;

Matches 478; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 1 CTTTGGTGTATAGACATGAAGTCTTGGCCATGCCCTATGCTCTGAATGGTATGGCCTG 60

|||||

Db 647 CTTTGGTGTATAGACATGAAGTCTTGGCCGCTATGCTCTGAATGGTATGGCCTA 588

|||||

QY 61 GGTTCCTCTAGGATTTATAGTTTATAGTCTAAACATTTAAGTCTTTAATCCATCTTG 120

|||||

Db 587 GGTTCCTCTAGGATTTATAGTTTATAGTCTAAACGTTTAAAGTGTTTAATCCATCTTG 528

|||||

QY 121 AATTAATTTTGTATAGTCTAAGGATGAAGAGGATCCAGTTTCAGCTTCTACATAGGCCTA 180

|||||

Db 527 AATTAATTTTGTATAGTCTAAGGATGAAGAGGATCCAGTTTCAGCTTCTACATAGGCCTA 468

|||||

QY 181 GCCAGTTTCTCAGCACCATTATTAATAGGGAATCCTTTCCCATTCGTTGTTTCT 240

|||||

Db 467 GCCAGTTTCTCAGCACCATTATTAATAGGGAATCCTTTCCCATTCGTTGTTTCT 408

|||||

QY 241 CAGGTTTCTCAAGATCAGATAGTTGTAGATATGCGCATTTATTTCTGAGGCTCTGTTTC 300

|||||

Db 407 CAGGTTTCTCAAGATCAGATAGTTGTAGATATGCGCGTTATTTCTGAGGCTCTGTTTC 348

|||||

QY 301 TGTTCATTCGTGATATCTCTGTTTGGTACCATCATCTGTTTGGTACTGTAG 360

|||||

Db 347 TGTTCATTCGTGATATCTCTGTTTGGTACCATCATCTGTTTGGTACTGTAG 288

|||||

QY 361 CCTTGTAGTGTAGTTGAAGTCAGGTAGCATGATGCCCTCCAGCTTTGTTCTTTTGGCTTA 420

|||||

Db 287 CCTTGTAGTGTAGTTGAAGTCAGGTAGCATGATGCCCTCCAGCTTTGTTCTTTTGGCTTA 228

|||||

QY 421 GGATTGACTTGGCATGTGGGCTCTTTTGGTTCACATGAACCTTTAAAGTAGTTTTTT 480

|||||

Db 227 GGATTGACTTGGCATGTGGGCTCTTTTGGTTCACATGAACCTTTAAAGTAGTTTTTT 168

|||||

QY 481 CCAATTCGTGAAGAAAGTAA 501

|||||

Db 167 CCAATTCGTGAAGAAAGTCA 147

|||||

RESULT 7

BM549258

LOCUS

DEFINITION BM549258 1069 bp mRNA linear EST 20-FEB-2002

5', mRNA sequence. AGENCOURT_6558415 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5548510

ACCESSION BM549258

VERSION BM549258.1

KEYWORDS EST. GI:18784517

SOURCE human.

ORGANISM Homo sapiens

REFERENCE

1 (bases 1 to 1069)

NIH-MGC <http://mgi.nci.nih.gov/>.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-re@mail.nih.gov

Tissue Procurement: ATCC/DCTD/DTF

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: L1AM12257 row: i column: 23

High quality sequence stop: 695.

FEATURES

source

Location/Qualifiers

1. 1069

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:5548510"

/clone_lib="NIH_MGC_72"

/tissue_type="melanotic melanoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;

Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.

Average insert size 2 kb. Library constructed by Life

Technologies."

BASE COUNT 236 a 193 c 224 g 412 t 4 others

ORIGIN

Query Match 92.7%; Score 464.2; DB 13; Length 1069;

Best Local Similarity 95.4%; Pred. No. 3.1e-104;

Matches 478; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 1 CTTTGGTGTATAGACATGAAGTCTTGGCCATGCCCTATGCTCTGAATGGTATGGCCTG 60

Db 80 CTTTGGTGTATAGACATGAAGTCTTGGCCATGCCCTATGCTCTGAATGGTATGGCCTA 139

QY 61 GGTTCCTCTAGGATTTATAGTTTATAGTCTAAACATTTAAGTCTTTAATCCATCTTG 120

Db 140 GGTTCCTCTAGGATTTATAGTTTATAGTCTAAACATTTAAGTCTTTAATCCATCTTG 199

QY 121 AATTAATTTTGTATAGTCTAAGGATGAAGAGGATCCAGTTTTCAGCTTCTACATAGGCCTA 180

Db 200 AATTAATTTTGTATAGGTTGAAGAGGATCCAGTTTTCAGCTTCTACATAGGCCTA 259

QY 181 GCCAGTTTCTCAGCACCATTATTAATAGGGAATCCTTTCCCATTCGTTGTTTCT 240

Db 260 GCCAGTTTCTCAGCACCATTATTAATAGGGAATCCTTTCCCATTCGTTGTTTCT 319

QY 241 CAGGTTTCTCAAGATCAGATAGTTGTAGATATGCGCATTTATTTCTGAGGCTCTGTTTC 300

Db 320 CAGGTTTCTCAAGATCAGATAGTTGTAGATATGCGCATTTATTTCTGAGGCTCTGTTTC 379

QY 301 TGTTCATTCGTGATATCTCTGTTTGGTACCATCATCTGTTTGGTACTGTAG 360

Db 380 TGTTCATTCGTGATATCTCTGTTTGGTACCATCATCTGTTTGGTACTGTAG 439

QY 361 CTTGTAGTGTAGTTGAAGTCAGGTAGCATGATGCCCTCCAGCTTTGTTGTTTGGCTTA 420

Db 440 CTTGTAGTGTAGTTGAAGTCAGGTAGCATGATGCCCTCCAGCTTTGTTGTTTGGCTTA 499

QY 421 GGATTGACTTGGCATGTGGGCTCTTTTGGTTCACATGAACCTTTAAAGTAGTTTTTT 480

Db 500 GGATTGACTTGGCATGTGGGCTCTTTTGGTTCACATGAACCTTTAAAGTAGTTTTTT 559

QY 481 CCAATTCGTGAAGAAAGTAA 501

Db 560 CCAATTCGTGAAGAAAGTCA 580

RESULT 8

AG102893 673 bp DNA linear GSS 03-NOV-2001
 LOCUS Pan troglodytes DNA, clone: PTB-106F11.F, genomic survey sequence.
 DEFINITION
 AG102893
 ACCESSION
 AG102893.1 GI:16723410
 VERSION
 GSS.
 KEYWORDS
 SOURCE

ORGANISM

Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
 BAC Library clone:PTB-106F11.F.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

REFERENCE

1 Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
 Totoki,Y., Watanabe,H. and Sakaki,Y.

AUTHORS

BAC end sequences of Library PTB

TITLE

Unpublished

JOURNAL

2 (bases 1 to 673)

REFERENCE

Totoki,Y., Watanabe,H. and Sakaki,Y.

AUTHORS

Direct Submission

TITLE

Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical

JOURNAL

and Chemical Research (RIKEN), Genomic Sciences Center (GSC);

1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

(E-mail:chmpbes@sc.riken.go.jp, URL:http://hgpc.gsc.riken.go.jp/,

Tel:81-45-503-9111, Fax:81-45-503-9170)

Clones are derived from the chimpanzee BAC library PTB This BAC end

was generated during the R&D process and may have higher chance of

clone tracking errors.

PRIMERS

SEQUENCING

-21M13

LIBRARY

Vector : pKS145

R.Site 1 : SacI

R.Site 2 : SacI

Location/Qualifiers

1. 673

/organism="Pan troglodytes"

/db_xref="taxon:9598"

/clone="PTB-106F11.F"

/sex="male"

/cell_type="lymphoblast"

/clone_lib="PTB Chimpanzee Male BAC Library"

137 a 119 c 152 g 262 t 3 others

BASE COUNT

ORIGIN

Query Match 92.3%; Score 462.6; DB 17; Length 673;

Best Local Similarity 95.2%; Pred. No. 6.7e-104;

Mismatches 0; Mismatches 24; Indels 0; Gaps 0;

Matches 477; Conservative 0;

1 CTTTGGTCTTTAGACATGAAGTCTTGCCTATGCTGCTGAATGGTATGCTG 60

136 CTTTGGTCTTTAGACATGAAGTCTTGCCTATGCTGCTGAATGGTATGCTG 195

61 GGTCTTCTAGGATTTTATGGTTTATAGTCTAACATTAAAGCTTTTAATCCATCTG 120

196 GGTCTTCTAGGATTTTATGGTTTATAGTCTAACATTAAAGCTTTTAATCCATCTG 255

121 AATTAATTTTGTATAGCTGTAAGGAGGATCCAGTTTCAGCTTCTACATAGGCTA 180

256 AATTAATTTTGTATAGCTGTAAGGAGGATCCAGTTTCAGCTTCTACATAGGCTA 315

181 GCCAGTTTCTCAGCACCATTATTAATAGGGAATCCCTTCCCATTCCTGTTTCT 240

316 GCCAGTTTCTCAGCACCATTATTAATAGGGAATCCCTTCCCATTCCTGTTTCT 375

241 CAGTTTGTCAAGATCAGATAGTGTAGATAGGGCAATTAATTCAGAGGCTCTGTC 300

376 CAGTTTGTCAAGATCAGATAGTGTAGATAGGGCAATTAATTCAGAGGCTCTGTC 435

301 TGTTCATGGTTGATATCTCTGTTTGGTACCAGTACCACATCTGTTTGGTACTGTAG 360

436 TGTTCATGGTTGATATCTCTGTTTGGTACCAGTACCACATCTGTTTGGTACTGTAG 495

FEATURES

source

QY 361 CCTTGTAGTCTAGTTTAAAGTCAGGTAGCATGATGCTCCAGCTTGTCTTTTGGCTTA 420
 Db 496 CCTTGTAGTCTAGTTTAAAGTCAGGTAGCATGATGCTCCCGCTTGTCTTTTGGCTTA 555
 QY 421 GGATTGACATTGGCGATGTGGCTCTTTTGGTTCCACATGAACACTTAAAGTAGTATTTT 480
 Db 556 GGATTGACATTGGCAATGCGAGGCTCTTTTGGTTCCATATGAACACTTAAAGTAGTATTTT 615
 QY 481 CCAATCTCTGGAAGAAATAA 501
 Db 616 CCAATCTCTGGAAGAAATAA 636

RESULT 9

AW850541/C

LOCUS

DEFINITION

IL3-CT0219-280100-062-F07 CT0219 Homo sapiens cDNA, mRNA sequence.

AW850541

ACCESSION

VERSION

AW850541.1 GI:7946058

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 681)

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,

Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,

Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,

Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare

,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and

Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

CONTACT: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6&2=IL3-CT0219-280

100-062-F07&t3=2000-01-28&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 5

High quality sequence stop: 535.

Location/Qualifiers

1. 681

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="CT0219"

/dev_stage="Adult"

/note="Organ: colon; Vector: puc18; Site 1: SmaI; Site 2:

SmaI; A mini-library was made by cloning products derived

from ORESTES PCR (U.S. Letters Patent application No. 196

,716 - Ludwig Institute for Cancer Research) profiles

into the puc 18 vector. Reverse transcription of tissue

mRNA and cDNA amplification were performed under low

stringency conditions."

BASE COUNT 270 a 148 c 122 g 141 t

ORIGIN

Query Match 92.0%; Score 461; DB 10; Length 681;

Best Local Similarity 95.0%; Pred. No. 1.7e-103;

Matches 476; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

1 CTTTGGTGTGTTTACACATGAAGTCTTGCCTATGCTGCTGAATGGTATGCTG 60

library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from
Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tldb/hungen/bac_end_search/bac_end_search.html
Seq primer: 77
Class: BAC ends.

FEATURES	source	Class: unc ends:	Location/Qualifiers
		1. .695	
		/organism="Homo sapiens"	
		/db_xref="GDB:7544347"	
		/db_xref="taxon:9606"	
		/clone="RPCI-11-116H20"	
		/clone_lib="RPCI-11"	
		/sex="Male"	
		/cell_type="Lymphocytes"	
		/note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI; RPCI11 Human Male BAC Library"	
BASE COUNT	285 a	153 c	118 g 139 t
ORIGIN			
		Query Match	91.7%; Score 459.4; DB 17; Length 695;
		Best Local Similarity	94.8%; Pred. No. 4.2e-103;
		Matches	475; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
QY	1	CTTTTGGTGTTTAGACATCAAGTCCCTTGCCCATGCCCTATGCTCCTGAATGGTATTGCTG	60
Db	654	CTTTGAGTGTTTTAGACATCAAGTCTTTGGCCCATGCCCTATGCTCCTGAATGGTAAATGCTTA	595
QY	61	GGTTTTCTTCTAGGGATTTTATGGTTTATAGTCTTAACATTTAAAGTCTTTTAAATCCATCTTG	120
Db	594	GGTTTCTTCTAGGGTTTTCATGGTTTATAGTCTTAAGTCTTTTAAATCCATCTTG	535
QY	121	AATTAATTTTGTATAGGTGTGAAGGAAGGATCCAGTTTCAGCTTTCTACATAGGCTA	180
Db	534	AATTGATTTTGTATAGGTGTGAAGGAAGGATCCAGTTTCAGCTTTCTCCATATGGCTA	475
QY	181	GCAGTTTTCTCAGCACCATTTATTAAATAGGGAATCCTTTCCCATGCTGTTGTTTCT	240
Db	474	GCAGTTTTCTCAGCACCATTTATTAAATAGGGAATCCTTTCCCATGCTGTTGTTTCT	415
QY	241	CAGGTTTGTCAAAGATCAGATAGTTGTAGATATGCGGCATATTCTCGAGGCTCTGTTCT	300
Db	414	CAGGTTTGTCAAAGATCAGATAGTTGTAGATATGCGGCGTTATTCTCGAGGCTCTGTTCT	355
QY	301	TGTTCCATGGTTGATATCTCTGTTTGGTACCAGTACCAGTACCATGTTGTTTGGTACTGTAG	360
Db	354	TGTTCCATGATCTATATCTCTGTTTGGTACCAGTACCAGTACCATGCTGTTTGGTACTGTAG	295
QY	361	CCTTGTAGTCTAGTTTGAAGTCAGGTAGCATGCTCCAGCTTTGTTCTTTTGGCTTA	420
Db	294	CCTTGTAGTATAGTTTGAAGTCAGGTAGTGTGATGCTCCAGCTTTGTTCTTTTGGCTTA	235
QY	421	GGATTGACTTGGCGATGCTGGCGCTCTTTTGGTTCCACATGAACCTTTAAAGTAGTGTTTT	480
Db	234	GGATTGACTTGGTGTATGATGAGGCGCTCTTTTGGTTCCATATGAACCTTTAAAGTAGTGTTTT	175
QY	481	CCAAATCTGTGAAGAAGTAA	501
Db	174	CCAAATCTGTGAAGAAGTCA	154

RESULT 12	
BQ071465/c	
LOCUS	BQ071465 linear EST 02-APR-2002
DEFINITION	AGENCOURT_5867166 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:5927882
	5' , mRNA sequence.
ACCESSION	BQ071465
VERSION	BQ071465.1 GI:19900511
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 1049)
NIH-MGC <http://mgc.nci.nih.gov/>,
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLCM2103 row: 1 column: 03
High quality sequence stop: 602.

FEATURES source

```

FEATURES
  source
    high_quality_sequence: 90p. 002.
    location/Qualifiers
      1. .1049
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="IMAGE:5927882"
      /clone_lib="NIH_MGC_47"
      /tissue_type="neuroblastoma, cell line"
      /lab_host="DH10B (phage-resistant)"
      /note="Organ: brain; Vector: pOTB7; Site:1: XhoI; Site:2:
      EcoRI; cDNA made by oligo-dT priming. Directionally
      cloned into EcoRI/XhoI sites using the following 5',
      adaptor: GGCACGAG(G). Size-selected >500bp for average
      insert size 1.8kb. Library constructed by Ling Hong in
      the laboratory of Gerald M. Rubin (University of
      California, Berkeley) using ZAP-cDNA synthesis kit
      (stratagene) and Superscript II RT (Life Technologies).
      Note: this is a NIH MGC Library."

```

	BASE COUNT	ORIGIN
1	0.0000	0.0000
2	0.0000	0.0000
3	0.0000	0.0000
4	0.0000	0.0000
5	0.0000	0.0000
6	0.0000	0.0000
7	0.0000	0.0000
8	0.0000	0.0000
9	0.0000	0.0000
10	0.0000	0.0000
11	0.0000	0.0000
12	0.0000	0.0000
13	0.0000	0.0000
14	0.0000	0.0000
15	0.0000	0.0000
16	0.0000	0.0000
17	0.0000	0.0000
18	0.0000	0.0000
19	0.0000	0.0000
20	0.0000	0.0000
21	0.0000	0.0000
22	0.0000	0.0000
23	0.0000	0.0000
24	0.0000	0.0000
25	0.0000	0.0000
26	0.0000	0.0000
27	0.0000	0.0000
28	0.0000	0.0000
29	0.0000	0.0000
30	0.0000	0.0000
31	0.0000	0.0000
32	0.0000	0.0000
33	0.0000	0.0000
34	0.0000	0.0000
35	0.0000	0.0000
36	0.0000	0.0000
37	0.0000	0.0000
38	0.0000	0.0000
39	0.0000	0.0000
40	0.0000	0.0000
41	0.0000	0.0000
42	0.0000	0.0000
43	0.0000	0.0000
44	0.0000	0.0000
45	0.0000	0.0000
46	0.0000	0.0000
47	0.0000	0.0000
48	0.0000	0.0000
49	0.0000	0.0000
50	0.0000	0.0000
51	0.0000	0.0000
52	0.0000	0.0000
53	0.0000	0.0000
54	0.0000	0.0000
55	0.0000	0.0000
56	0.0000	0.0000
57	0.0000	0.0000
58	0.0000	0.0000
59	0.0000	0.0000
60	0.0000	0.0000
61	0.0000	0.0000
62	0.0000	0.0000
63	0.0000	0.0000
64	0.0000	0.0000
65	0.0000	0.0000
66	0.0000	0.0000
67	0.0000	0.0000
68	0.0000	0.0000
69	0.0000	0.0000
70	0.0000	0.0000
71	0.0000	0.0000
72	0.0000	0.0000
73	0.0000	0.0000
74	0.0000	0.0000
75	0.0000	0.0000
76	0.0000	0.0000
77	0.0000	0.0000
78	0.0000	0.0000
79	0.0000	0.0000
80	0.0000	0.0000
81	0.0000	0.0000
82	0.0000	0.0000
83	0.0000	0.0000
84	0.0000	0.0000
85	0.0000	0.0000
86	0.0000	0.0000
87	0.0000	0.0000
88	0.0000	0.0000
89	0.0000	0.0000
90	0.0000	0.0000
91	0.0000	0.0000
92	0.0000	0.0000
93	0.0000	0.0000
94	0.0000	0.0000
95	0.0000	0.0000
96	0.0000	0.0000
97	0.0000	0.0000
98	0.0000	0.0000
99	0.0000	0.0000
100	0.0000	0.0000

Query Match 91.5%; Score 458.4; DB 14; Length 1049;
Best Local Similarity 94.6%; Pred. No. 8.2e-103;

60	QY	1	CTTTGGTGTTTTAGACATGAAGTCTCTTGCCCATGCTATGTCTCCTGAATGGTATATGCCG	60
721	Db			721
662	QY	61	GGTTTTCTCTAGGATTTTATGTGTTTTAGTCTCTAAACATTAAGTCTTTAAATCCATCTTG	662
602	Db	661	GGTTTTCTCTAGGTTTTTATGTTTTAGGTCCTAAGCTTAAAGTCTNTAAATCCCTTC	602
180	QY	121	AATTAATTTTTGTATAGGCTAAGGAAGGATCCAGTTTCAGCTTTCACATAGGGCTTA	180
542	Db	601	AATTAATTTTTGTATAGGCTAAGGAAGGATCCAGTTTCAGCTTTCACATAGGGCTTA	542
240	QY	181	GCGAGTTTCTCAGCACCATTTATTAATAGGGAATCCTTTCCCATTCGCTGTTTTTCT	240
482	Db	541	GCGAGTTTCCGAGCACCATTTATTAATAGGGAATCCTTTCCCATTCGCTGTTTTTCT	482
300	QY	241	CAGGTTTTCTCAAGATCAGATAGTGTAGATATCGGSCATTAATTTCTGAGGGCTCTGTTC	300
422	Db	481	CAGGTTTTCTCAAGATCAGATAGTGTAGATATCGGSCATTAATTTCTGAGGGCTCTGTTC	422
360	QY	301	TGTTCCATGGTTGATATCTCTGTTTTGGTACCAAGTACCAATGTTGTTTTGGTTACTGTAG	360
362	Db	421	TGTTCCATGGTTGATATCTCTGTTTTGGTACCAAGTACCAATGTTGTTTTGGTTACTGTAG	362
420	QY	361	CCTCTAGTGTAGTTTGAAGTCAGGTAGCATGATGCTCCAGCTTTGTTCTTTTTGGCTTA	420
302	Db	361	CCTCTAGTGTAGTTTGAAGTCAGGTAGCATGATGCTCCAGCTTTGTTCTTTTTGGCTTA	302
480	QY	421	GGATTGACTTGGCGATGGGCTCTTTTTGGTTCACATGAAGTAAAGTAGTTTTTT	480
242	Db	301	GGATTGACTTGGCGAATCGGGCTCTTTTTGGTTCACATGAAGTAAAGTAGTTTTTT	242
501	QY	481	CCAAATTCGTGAAGAAAGTAA 501	501

```

Db      241 CCAATTCTGTGAAGAAAGTCA 221
|||||
RESULT 13
LOCUS   BG288152/c
DEFINITION 602387957F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4516526 5',
mRNA sequence.
ACCESSION BG288152
VERSION   BG288152.1 GI:13042702
KEYWORDS EST.
SOURCE   human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 804)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: Incyte Genomics, Inc.
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10408 row: b column: 15
High quality sequence stop: 799.
FEATURES             source
Location/Qualifiers
     1..804
     /organism="Homo sapiens"
     /db_xref="taxon:9606"
     /clone_image="IMAGE:4516526"
     /clone_lib="NIH_MGC_93"
     /tissue_type="transitional cell papilloma, cell line"
     /lab_host="DH10B (phage-resistant)"
     /note="Organ: bladder; Vector: pCMV-SPORT6; Site_1: NotI;
     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
     Average insert size 1.7 kb. Library enriched for
     full-length clones and constructed by Life Technologies.
     Note: this is a NIH_MGC Library."
BASE COUNT 325 a 180 c 135 g 164 t
ORIGIN
Query Match 91.58; Score 458.2; DB 12; Length 804;
Best Local Similarity 96.28; Pred. No. 8.6e-103;
Matches 480; Conservative 0; Mismatches 18; Indels 1; Gaps 1;
Qy 1 CTTTGGTGTGTAGACATCAAGTCCTTGGCCATGCTATGCTCTGAATGGT-ATTCCCT 59
|||||
Db 615 CTTTGGTGTGTAGACATCAAGTCCTTGGCCATGCTATGCTCTGAATGGTAAATGCCT 556
|||||
Qy 60 GGGTTTCTCTAGGGATTTTATGGTTTATAGGCTCAACATTTAAGTCTTTAATCCATCTT 119
|||||
Db 555 AGGTTTCTCTAGGGTTTATGGTTTATAGGCTCAACGTTTAAAGTCTTTAATCCATCTT 496
|||||
Qy 120 GAATTAATTTTGTATTAAGCTGTAAAGGAAGGATCCAGTTTCAGCTTTTCATATAGGGCT 179
|||||
Db 495 GAATTTGATTTTGTATTAAGGTTAAAGGAAGGATCCAGTTTTCAGCTTTTCATATAGGCT 436
|||||
Qy 180 AGCCAGTTTTCAGCAGCAATTTATTAATAGGAATCCCTTCCCATTCCTGCTTTTTC 239
|||||
Db 435 AGCCAGTTTTCAGCAGCAATTTATTAATAGGAATCCCTTCCCATTCCTGCTTTTTC 376
|||||
Qy 240 TCAGGTTTGTCAAGATCAGATAGTTGTAGATATGCGGCATTTATTTCTGAGGGCTCTGTT 299
|||||
Db 375 TCAGGTTTGTCAAGATCAGATAGTTGTAGATATGCGGCCTTATTTCTGAGGGCTCTGTT 316
|||||
Qy 300 CTGTTCCATTTGGTTGATATCTCTGTTTGGTTACAGTACATGTTGTTTGGTTACTGTA 359
|||||

```

```

Db      315 CTGTTCCAGTGATCTATATCTCTCTTTTGGTACCAGTACCATGCTGTTTGGTTACTGTA 256
Qy 360 GCCTTGTAGTCTAGCTTTGAAGTCAGGTAGCATGCTCCAGCTTTGCTCTTTGGGCTT 419
|||||
Db 255 GCCTTGTAGTATATAGCTTTGAAGTCAGGTAGCTGATGCTCCAGCTTTGCTCTTTGGCTT 196
|||||
Qy 420 AGGATTGACCTGGCGATGTGGGCTCTTTTGGTTCCACATGAACCTTTAAAGTAGTATTTT 479
|||||
Db 195 AGGATTGACCTGGCGATATGGCTCTTTTGGTTCCATATGAACCTTTAAAGTAGTATTTT 136
|||||
Qy 480 TCCAAATTCCTGTGAAGAAAG 498
|||||
Db 135 TCCAAATTCCTGTGAAGAAAG 117
|||||
RESULT 14
LOCUS   AQ491435/c
DEFINITION RPCI-11-244B1.TV RPCI-11 Homo sapiens genomic clone RPCI-11-244B1,
DNA sequence.
ACCESSION AQ491435
VERSION   AQ491435.1 GI:4677309
KEYWORDS GSS.
SOURCE   human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 679)
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
,J.C.
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
Map Building
JOURNAL Unpublished (1997)
COMMENT Other_GSSs: RPCI-11-244B1.TJ
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genet cs (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: T7
Class: BAC ends.
FEATURES             source
Location/Qualifiers
     1..679
     /organism="Homo sapiens"
     /db_xref="GDB:7593336"
     /db_xref="taxon:9606"
     /clone_image="RPCI-11-244B1"
     /clone_lib="RPCI-11"
     /sex="Male"
     /cell_type="Lymphocytes"
     /note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
     RPCI11 Human Male BAC Library"
BASE COUNT 273 a 146 c 118 g 142 t
ORIGIN
Query Match 91.48; Score 457.8; DB 17; Length 679;
Best Local Similarity 94.68; Pred. No. 1e-102;
Matches 474; Conservative 0; Mismatches 27; Indels 0; Gaps 0;
Qy 1 CTTTGGTGTGTAGACATCAAGTCCTTGGCCATGCTATGCTCTGAATGGTATTTGCTG 60
|||||
Db 654 CTTTGGTGTGTAGACATCAAGTCCTTGGCCATGCTATGCTCTGAATGGTATTTGCTG 595
|||||
Qy 61 GGTTCCTCTAGGGATTTTATGGTTTATAGTCTTAACATTTAAGTCTTTAATCCATCTG 120
|||||

```

```
Db 594 GGTTCCTCTAGGGTTTTATGGTTTAAAGTCTAACATGTAAGTCTTTAATCCACCTG 535
Qy 121 AATTAATTTTGTATAAGGTGTAAGGAAGGATCCAGTTTTCAGCTTTTACATAGGCTA 180
Db 534 AATTAATTTTGTATAAGGTGTAAGGAAGGATACAGTTTAACTTTTACATATGGCTA 475
Qy 181 GCCAGTTTCTCAGCACCATTATTAATAAGGAAGTCCCTTCCCATATGCTTTTCT 240
Db 474 GCCAGTTTCCAGCACCATTATTAATAAGGAAGTCCCTTCCCATATGCTTTTCT 415
Qy 241 CAGTTTGTCAAAGATCAGATAGTGTAGATATGCGCATTAATTTCTGAGGGCTCTGTT 300
Db 414 CAGTTTGTCAAAGATCAGATAGTGTAGATATGCGCATTAATTTCTGAGGGCTCTGTT 355
Qy 301 TGTTCATGGTTGTATCTCTCTTTTGGTACCAGTACCAGTGTGTTTGGTTTACGTAG 360
Db 354 TGTTCATGGTTGTATCTCTCTTTTGGTACCAGTACCAGTGTGTTTGGTTTACGTAG 295
Qy 361 CCTGTAGTGTAGTTTGAAGTACAGTACATGATGCTCCAGCTTTTGTCTTTTGGCTTA 420
Db 294 CCTGTAGTGTAGTTTGAAGTACAGTACATGATGCTCCAGCTTTTGTCTTTTGGCTTA 235
Qy 421 GGATTGACTTGGGATGTGGGCTCTTTTGGTTCACATGAACCTTAAAGTAGTTT 480
Db 234 GGATTGACTTGGCAATGCAAGGCTCTTTTGGTTCACATGAACCTTAAAGTAGTTT 175
Qy 481 CCAATTCCTGTGAAGAAAGTAA 501
Db 174 CCAATTCCTGTGAAGAAAGTCA 154

RESULT 15
AG077683
LOCUS Pan troglodytes DNA, clone: PTB-072E06.F, linear GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-072E06.F, genomic survey sequence.
ACCESSION AG077683
VERSION AG077683.1 GI:16629485
KEYWORDS GSS.
SOURCE Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
BAC Library clone:PTB-072E06.F.
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totohi,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of Library PTB
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 656)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totohi,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chmpes@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI
Location/Qualifiers
1. .656
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-072E06.F"
/sex="male"
/cell_type="lymphoblast"
```

```
BASE COUNT 131 a 118 c 153 g 254 t
ORIGIN
Query Match 91.3%; Score 457.2; DB 17; Length 656;
Best Local Similarity 96.2%; Pred. No. 1.4e-102; Indels 1; Gaps 1;
Matches 479; Conservative 0; Mismatches 18;
Qy 1 CTTTTCGTGTTTAGACATGAAGTCCCTTGCCTATGCTCCTGAATGGTATGCTG 60
Db 131 CTTTTCGTGTTTAGACATGAAGTCCCTTGCCTATGCTCCTGAATGGTATGCTG 190
Qy 61 GGTTCCTCTAGGATTTTATGGTTTATAGTCTAATTAAGTCTTAAATCCATCTG 120
Db 191 GGTTCCTCTAGGATTTTATGGTTTATAGTCTAATTAAGTCTTAAATCCATCTG 250
Qy 121 AATTAATTTTGTATAAGGTGTAAGGAAGGATCCAGTTTTCAGCTTTTACATAGGCTA 180
Db 251 AATTAATTTTGTATAAGGTGTAAGGAAGGATCCAGTTTTCAGCTTTTACATAGGCTA 309
Qy 181 GCCAGTTTCTCAGCACCATTATTAATAAGGAATCCCTTCCCATATGCTTTTCT 240
Db 310 GCCAGTTTCCAGCACCATTATTAATAAGGAATCCCTTCCCATATGCTTTTCT 369
Qy 241 CAGTTTGTCAAAGATCAGATAGTGTAGATATGCGCATTAATTTCTGAGGGCTCTGTTTC 300
Db 370 CAGTTTGTCAAAGATCAGATAGTGTAGATATGCGGGCTTAATTTCTGAGGGCTCTGTTTC 429
Qy 301 TGTTCATTTGGTGTATCTCTGTTTGGTACCAGTACCAGTGTGTTTGGTACGTAG 360
Db 430 TGTTCATTTGGTGTATCTCTGTTTGGTACCAGTACCAGTGTGTTTGGTACGTAG 489
Qy 361 CCTTGTAGTGTAGTTTGAAGTCAAGTCAAGTCAAGTCTTCTTTTGGCTTA 420
Db 490 CCTTGTAGTGTAGTTTGAAGTCAAGTCAAGTCAAGTCTTCTTTTGGCTTA 549
Qy 421 GGATTGACTTGGGATGTGGGCTCTTTTGGTTCACATGAACCTTAAAGTAGTTT 480
Db 550 GGATTGACTTGGGATGTGGGCTCTTTTGGTTCACATGAACCTTAAAGTAGTTT 609
Qy 481 CCAATTCCTGTGAAGAAAG 498
Db 610 CCAATTCCTGTGAAGAAAG 627

Search completed: July 4, 2003, 17:35:43
Job time : 1099.73 secs
```

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	501	100.0	29921	9	US-10-083-853-1		Sequence 1, Appli
2	478.6	95.5	5065	9	US-10-050-882-39		Sequence 39, Appl
C	477.4	95.3	4709	9	US-10-050-882-40		Sequence 40, Appl
4	472.8	94.4	326014	10	US-09-771-231A-3		Sequence 3, Appl
C	472.2	94.3	169327	10	US-09-777-921A-3		Sequence 3, Appl
C	472.2	94.3	167343	10	US-09-962-436-281		Sequence 281, Appl
C	472.2	94.3	167343	10	US-09-964-824A-273		Sequence 273, Appl
8	469	93.6	180557	12	US-10-003-806-5		Sequence 6, Appl
C	467.4	93.3	16884	9	US-09-764-891-8497		Sequence 9, Appl
11	467.4	93.3	32042	9	US-10-118-984-44		Sequence 44, Appl
12	467.4	93.3	32042	9	US-10-295-981-63		Sequence 63, Appl
13	467.4	93.3	32042	10	US-09-728-721-63		Sequence 63, App
14	467.4	93.3	34001	9	US-10-006-883A-15		Sequence 15, Appl
C	467.4	93.3	1691139	9	US-10-067-514-1		Sequence 1, Appl
16	466	93.0	11057	9	US-09-764-868-1308		Sequence 1308, App
17	465.8	93.0	6175	9	US-10-079-854-294		Sequence 294, App
C	465.8	93.0	6175	10	US-09-764-878-294		Sequence 294, App
C	465.8	93.0	7626	9	US-10-001-835-82		Sequence 82, Appl

QY 241 CAGTTTGTCAAAGATCAGATAGTTGTAGATATCGGCATTAATTTCTGAGGGCTCTGTTTC 300
Db 29661 CAGTTTGTCAAAGATCAGATAGTTGTAGATATCGGCATTAATTTCTGAGGGCTCTGTTTC 29720
QY 301 TGTTCATTTGGTTGATATCTCTGTTTGGTACCAGTACCAGTTGTTTGGTTACTGTAG 360
Db 29721 TGTTCATTTGGTTGATATCTCTGTTTGGTACCAGTACCAGTTGTTTGGTTACTGTAG 29780
QY 361 CTTTGTAGTGTAGTTTGAAGTCAGGTAGCATGATGCTCCAGCTTTCTTTTGGGCTTA 420
Db 29781 CTTTGTAGTGTAGTTTGAAGTCAGGTAGCATGATGCTCCAGCTTTCTTTTGGGCTTA 29840
QY 421 GGATTGACTTGGCGATGCGGCTCTTTTGGTCCACATGAACCTTTAAAGTAGTTTTTT 480
Db 29841 GGATTGACTTGGCGATGCGGCTCTTTTGGTCCACATGAACCTTTAAAGTAGTTTTTT 29900
QY 481 CCAATTCTGTGAAGAAAGTAA 501
Db 29901 CCAATTCTGTGAAGAAAGTAA 29921

RESULT 2

US-10-050-882-39

; Sequence 39, Application US/10050882

; Publication No. US20030104400A1

; GENERAL INFORMATION:

; APPLICANT: Ruben et al.

; TITLE OF INVENTION: 27 Human secreted proteins

; FILE REFERENCE: P2038P1

; CURRENT APPLICATION NUMBER: US/10/050,882

; CURRENT FILING DATE: 2002-01-18

; PRIOR FILING DATE: 2000-09-13

; PRIOR APPLICATION NUMBER: PCT/US00/06783

; PRIOR FILING DATE: 2000-03-16

; PRIOR APPLICATION NUMBER: 60/125,055

; PRIOR FILING DATE: 1999-03-18

; NUMBER OF SEQ ID NOS: 156

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 39

; LENGTH: 5065

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: SITE

; LOCATION: (2531)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: SITE

; LOCATION: (5063)

; OTHER INFORMATION: n equals a,t,g, or c

US-10-050-882-39

Query Match 95.5%; Score 478.6; DB 9; Length 5065;

Best Local Similarity 95.4%; Pred. No. 8.7e-123;

Matches 476; Conservative 11; Mismatches 11; Indels 0; Gaps 0;

QY 1 CTTTGTGTGTAGACATGAAGTCCTTGCCCATGCTATGCTCTGTAATGGTATTGCGCTG 60

Db 1090 CTTTGTGTGTAGACATGAAGTCCTTGCCCATGCTATGCTCTGTAATGGTATTGCGCTA 1149

QY 61 GGTTTCTTCTAGGATTTTATGGTTTATAGTCTAACATTTAAGTCTTTAATCCATCTTG 120

Db 1150 GGTTTCTTCTAGGATTTTATGGTTTATAGTCTAACATTTAAGTCTTTAATCCATCTTG 1209

QY 121 AATTAATTTTGTATAGGTGAAGGAGGATCCAGTTTCAGCTTCTACATAGGCTTA 180

Db 1210 AATTAATTTTGTATAGGTGAAGGAGGATCCAGTTTCAGCTTCTACATAGGCTTA 1269

QY 181 GCCAGTTTCTCAGCACCATTATTAATAAGGAATCCTTTCCCATGCTTCTGTTTCT 240

Db 1270 GCCAGTTTCTCAGCACCATTATTAATAAGGAATCCTTTCCCATGCTTCTGTTTCT 1329

QY 241 CAGTTTGTCAAAGATCAGATAGTTGTAGATATCGGCATTAATTTCTGAGGGCTCTGTTTC 300

Db 1330 CAGTTTGTCAAAGATCAGATAGTTGTAGATATGCGGCTCTGTTTC 1389
QY 301 TGTTCATTTGGTTGATATCTCTGTTTGGTACCAGTACCAGTTGTTTGGTTACTGTAG 360
Db 1390 TGTTCATTTGGTTGATATCTCTGTTTGGTACCAGTACCAGTTGTTTGGTTACTGTAG 1449
QY 361 CTTTGTAGTGTAGTTTGAAGTCAGGTAGCATGATGCTCCAGCTTTCTTTTGGGCTTA 420
Db 1450 CTTTGTAGTGTAGTTTGAAGTCAGGTAGCATGATGCTCCAGCTTTCTTTTGGGCTTA 1509
QY 421 GGATTGACTTGGCGATGCGGCTCTTTTGGTCCACATGAACCTTTAAAGTAGTTTTTT 480
Db 1510 GGATTGACTTGGCGATGCGGCTCTTTTGGTCCACATGAACCTTTAAAGTAGTTTTTT 1569
QY 481 CCAATTCTGTGAAGAAAGTAA 501
Db 1570 CCAATTCTGTGAAGAAAGTAA 1590

RESULT 3

US-10-050-882-40/c

; Sequence 40, Application US/10050882

; Publication No. US20030104400A1

; GENERAL INFORMATION:

; APPLICANT: Ruben et al.

; TITLE OF INVENTION: 27 Human secreted proteins

; FILE REFERENCE: P2038P1

; CURRENT APPLICATION NUMBER: US/10/050,882

; CURRENT FILING DATE: 2002-01-18

; PRIOR FILING DATE: 2000-09-13

; PRIOR APPLICATION NUMBER: PCT/US00/06783

; PRIOR FILING DATE: 2000-03-16

; PRIOR APPLICATION NUMBER: 60/125,055

; PRIOR FILING DATE: 1999-03-18

; NUMBER OF SEQ ID NOS: 156

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 40

; LENGTH: 4709

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: SITE

; LOCATION: (14)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: SITE

; LOCATION: (18)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: SITE

; LOCATION: (30)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: SITE

; LOCATION: (34)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: SITE

; LOCATION: (177)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: SITE

; LOCATION: (2213)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: SITE

; LOCATION: (3846)

; OTHER INFORMATION: n equals a,t,g, or c

US-10-050-882-40

Query Match 95.3%; Score 477.4; DB 9; Length 4709;

Best Local Similarity 95.4%; Pred. No. 1.8e-122;

Matches 478; Conservative 11; Mismatches 12; Indels 0; Gaps 0;

QY 1 CTTTGTGTGTAGACATGAAGTCCTTGCCCATGCTATGCTCTGAATGGTATTGCGCTG 60

Db 3654 CTTTGTGTGTGTAGACATGAAGTCCTTGCCCATGCTATGCTCTGAATGGTATTGCGCTA 3595

QY 61 GGTTCCTCTAGGGATTTATGGTTTATAGTCTAAACATTTAAGTCTTTAATCAATCTTG 120
|||||
Db 3594 GGTTCCTCTAGGGTTTATGGTTTATAGTCTAAACATTTAAGTCTTTAATCAATCTTG 3535

QY 121 AATTAATTTTGTATAGGTGTAAGGAAGGATCCAGTTTCAGCTTCTACATAGGCTA 180
|||||
Db 3534 AATTAATTTTGTATAGGTGTAAGGAAGGATCCAGTTTCAGCTTCTACATAGGCTA 3475

QY 181 GCCAGTTTCTCAGCACCACTTTATTAATAGGAATCCTTTCCCAATTCCTCTTTTCT 240
|||||
Db 3474 GCCAGTTTCCACAGCACCACTTTATTAATAGGAATCCTTTCCCAATTCCTCTTTTCT 3415

QY 241 CAGGTTTGTCAAAGATCAGATAGTCTGTAGATATGCGGCATTAATTTCTGAGGGCTCTGTTT 300
|||||
Db 3414 CAGGTTTGTCAAAGATCAGATAGTCTGTAGATATGCGGCATTAATTTCTGAGGGCTCTGTTT 3355

QY 301 TGTTCATCGTTGATATCTCTCTTTTGGTACCAGTACCAGTGTGTTGGTTACTGTAG 360
|||||
Db 3354 TGTTCATCGTTGATATCTCTCTTTTGGTACCAGTACCAGTGTGTTGGTTACTGTAG 3295

QY 361 CCTGTAGTGTAGTCTGAGTCAAGTATGATGCTCCAGCTTTGTTCTTTTGGCTTA 420
|||||
Db 3294 CCTGTAGTGTAGTCTGAGTCAAGTATGATGCTCCAGCTTTGTTCTTTTGGCTTA 3235

QY 421 GGATTCACCTGGCGATGTGGGCTCTTTTGGTTCCACATGAACCTTTAAAGTAGTTTTT 480
|||||
Db 3234 GGATTCACCTGGCGATGTGGGCTCTTTTGGTTCCACATGAACCTTTAAAGTAGTTTTT 3175

QY 481 CCAATTCCTGTGAAGAAAGTAA 501
|||||
Db 3174 CCAATTCCTGTGAAGAAAGTAA 3154

RESULT 4

US-09-731-231A-3

; Sequence 3, Application US/09731231A

; Patent No. US20020082189A1

; GENERAL INFORMATION:

; APPLICANT: MERKULOV et al.

; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC

; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES

; FILE REFERENCE: THEREOF

; FILE REFERENCE: CL001007

; CURRENT APPLICATION NUMBER: US/09/731,231A

; CURRENT FILING DATE: 2000-12-07

; NUMBER OF SEQ ID NOS: 6

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 3

; LENGTH: 326014

; TYPE: DNA

; ORGANISM: Human

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: (1)...(326014)

; OTHER INFORMATION: n = A,T,C or G

US-09-731-231A-3

Query Match 94.4%; Score 472.8; DB 10; Length 326014;
Best Local Similarity 96.8%; Pred. No. 2.7e-120;
Matches 483; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 2 TTTTGGTGTATAGCATGAAGTCCCTTGCCATGCTATGCTCTGAATGGTATTGCTGG 61
|||||
Db 31241 TTTTGGTGTATAGCATGAAGTCCCTTGCCATGCTATGCTCTGAATGGTATTGCTGG 31300

QY 62 GTTTCCTCTAGGGATTTTATGGTTTATAGTCTTAACATTTAAGTCTTTAATCCATCTGA 121
|||||
Db 31301 GTTTCCTCTAGGGATTTTATGGTTTATAGTCTTAACATTTAAGTCTTTAATCCATCTGA 31360

QY 122 ATTAATTTTGTATAGGTGTATAGGAAGGATCCAGTTTCAGCTTTCTACATAGGGCTAG 181
|||||
Db 31361 ATTAATTTTGTATAGGTGTATAGGAAGGATCCAGTTTCAGCTTTCTACATAGGGCTAG 31420

QY 182 CCAGTTTCTCAGCACCACTTTAATTAATAGGAATCCTTTCCCAATTCCTGTTTCTC 241
|||||
Db 31421 CCAGTTTCTCAGCACCACTTTAATTAATAGGAATCCTTTCCCAATTCCTGTTTCTC 31480

QY 242 AGTTTGTCAAAGATCAGATAGTGTAGATATCGGCATTAATTTCTGAGGGCTCTGTCT 301
|||||
Db 31481 AGTTTGTCAAAGATCAGATAGTGTAGATATCGGCATTAATTTCTGAGGGCTCTGTCT 31540

QY 302 GTTCCATTTGTTGATATCTCTCTTTTGGTACCAGTACCAGTGTGTTGGTTACTGTAG 361
|||||
Db 31541 GTTCCATTTGTTGATATCTCTCTTTTGGTACCAGTACCAGTGTGTTGGTTACTGTAG 31600

QY 362 CTTGTAGTGTAGTTTGAAGTCAAGTATGATGCTCCAGCTTTGTTCTTTTGGCTTAG 421
|||||
Db 31601 CTTGTAGTGTAGTTTGAAGTCAAGTATGATGCTCCAGCTTTGTTCTTTTGGCTTAG 31660

QY 422 GATTGACTTGGCGATGTGGGCTCTTTTGGTTCCACATGAACCTTTAAAGTAGTTTTTC 481
|||||
Db 31661 GATTGACTTGGCGATGTGGGCTCTTTTGGTTCCACATGAACCTTTAAAGTAGTTTTTC 31720

QY 482 CAATTCCTGTGAAGAAAGTAA 501
|||||
Db 31721 CAATTCCTGTGAAGAAAGTAA 31740

RESULT 5

US-09-777-921A-3/C

; Sequence 3, Application US/09777921A

; Patent No. US20020115136A1

; GENERAL INFORMATION:

; APPLICANT: MERKULOV et al.

; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,

; TITLE OF INVENTION: AND USES THEREOF

; FILE REFERENCE: CL001103

; CURRENT APPLICATION NUMBER: US/09/777,921A

; CURRENT FILING DATE: 2002-02-07

; NUMBER OF SEQ ID NOS: 126

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 3

; LENGTH: 69327

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: (1)...(69327)

; OTHER INFORMATION: n = A,T,C or G

US-09-777-921A-3

Query Match 94.3%; Score 472.2; DB 10; Length 69327;
Best Local Similarity 96.4%; Pred. No. 1.9e-120;
Matches 483; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 1 CTTTGGTGTATAGCATGAAGTCCCTTGCCATGCTATGCTCTGAATGGTATTGCTG 60
|||||
Db 40174 CTTTGGTGTATAGCATGAAGTCCCTTGCCATGCTATGCTCTGAATGGTATTGCTG 40115

QY 61 GGTTCCTCTAGGGATTTTATGGTTTATAGTCTTAACATTTAAGTCTTTAATCCATCTG 120
|||||
Db 40114 GGTTCCTCTAGGGATTTTATGGTTTATAGTCTTAACATTTAAGTCTTTAATCCATCTG 40055

QY 121 AATTAATTTTGTATAGGTGTATAGGAAGGATCCAGTTTCAGCTTTCTACATAGGCTA 180
|||||
Db 40054 AATTAATTTTGTATAGGTGTATAGGAAGGATCCAGTTTCAGCTTTCTACATAGGCTA 39995

QY 181 GCCAGTTTCTCAGCACCACTTTAATTAATAGGAATCCTTTCCCAATTCCTGTTTCT 240
|||||
Db 39994 GCCAGTTTCTCAGCACCACTTTAATTAATAGGAATCCTTTCCCAATTCCTGTTTCT 39935

QY 241 CAGGTTTGTCAAAGATCAGATAGTGTAGATATGCGGCATTAATTTCTGAGGGCTCTGTTC 300
|||||
Db 39934 CAGGTTTGTCAAAGATCAGATAGTGTAGATATGCGGCATTAATTTCTGAGGGCTCTGTTC 39875

QY 301 TGTTCATTTGGTGTATATCTCTGTTTGGTACCAGTACCAGTGTGTTTGGTTACTGTAG 360
|||||
Db 39874 TGTTCATTTGGTGTATATCTCTGTTTGGTACCAGTACCAGTGTGTTTGGTTACTGTAG 39815
QY 361 CTTGTAGTGTAGTTGAAGTCAGGTACATGATGCCCTCCAGCTTTGTTCTTTGGGCTTA 420
|||||
Db 39814 CTTGTAGTGTAGTTGAAGTCAGGTACATGATGCCCTCCAGCTTTGTTCTTTGGGCTTA 39755
QY 421 GGATTGACTTGGCGATGGGCTCTTTTGGTTCACATGAACCTTTAAAGTAGTTTTT 480
|||||
Db 39754 GGATTGACTTGGCAATGGGCTCTTTTGGTTCACATGAACCTTTAAAGTAGTTTTT 39695
QY 481 CCAATTCTGTGAAGAAAGTAA 501
|||||
Db 39694 CCAATTCTGTGAAGAAAGTCA 39674

RESULT 6
US-09-962-436-281/C
; Sequence 281, Application US/09962436
; Patent No. US20020081301A1
; GENERAL INFORMATION:
; APPLICANT: Soppec, Daniel
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
; FILE REFERENCE: 689290-75
; CURRENT APPLICATION NUMBER: US/09/962,436
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/60/235,082
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/234,924
; PRIOR FILING DATE: 2000-09-25
; NUMBER OF SEQ ID NOS: 568
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO: 281
; LENGTH: 167343
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-962-436-281

Query Match 94.3%; Score 472.2; DB 10; Length 167343;
Best Local Similarity 96.4%; Pred. No. 2.9e-120;
Matches 483; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
QY 1 CTTTGGTGTATAGACATGAAGTCCCTTGCCCATGCCCTATGCTCCTGAATGCTATTGCGTG 60
|||||
Db 59330 CTTTGGTGTATAGACATGAAGTCCCTTGCCCATGCCCTATGCTCCTGAAGGTAATGCCCTA 59271
QY 61 GGTTCCTCTAGGATTTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTT 120
|||||
Db 59270 GGTTCCTCTAGGATTTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTT 59211
QY 121 AATTAATTTTGTATAGGCTAAGGATCCAGTTTCAAGCTTTTAACTTTTAACTTTTAACTTT 180
|||||
Db 59210 AATTAATTTTGTATAGGCTAAGGATCCAGTTTCAAGCTTTTAACTTTTAACTTTTAACTTT 59151
QY 181 GCCAGTTTCTCAGCACCATTTATTAATAGGGAATCCTTTCCCATGCTTTGTTTCT 240
|||||
Db 59150 GCCAGTTTCTCAGCACCATTTATTAATAGGGAATCCTTTCCCATGCTTTGTTTCT 59091
QY 241 CAGGTTTGTCAAAGATCAGATAGTTAGATATCGGGCATTTATTTCTGAGGCTCTGTTTC 300
|||||
Db 59090 CAGGTTTGTCAAAGATCAGATAGTTAGATATCGGGCATTTATTTCTGAGGCTCTGTTTC 59031
QY 301 TGTTCATTTGGTGTATATCTCTGTTTGGTACCAGTACCAGTGTGTTTGGTTACTGTAG 360
|||||
Db 59030 TGTTCATTTGGTGTATATCTCTGTTTGGTACCAGTACCAGTGTGTTTGGTTACTGTAG 58971
QY 361 CTTGTAGTGTAGTTGAAGTCAGGTACATGATGCCCTCCAGCTTTGTTCTTTGGGCTTA 420
|||||
Db 58970 CTTGTAGTGTAGTTGAAGTCAGGTACATGATGCCCTCCAGCTTTGTTCTTTGGGCTTA 58911

QY 421 GGATTGACTTGGCGATGGGCTCTTTTGGTTCACATGAACCTTTAAAGTAGTTTTT 480
|||||
Db 58910 GGATTGACTTGGCAATGGGCTCTTTTGGTTCACATGAACCTTTAAAGTAGTTTTT 58851
QY 481 CCAATTCTGTGAAGAAAGTAA 501
|||||
Db 58850 CCAATTCTGTGAAGAAAGTCA 58830
RESULT 7
US-09-964-824A-273/C
; Sequence 273, Application US/09964824A
; Patent No. US20020102531A1
; GENERAL INFORMATION:
; APPLICANT: Horiigan, Stephen
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Sign
; FILE REFERENCE: 689290-73
; CURRENT APPLICATION NUMBER: US/09/964,824A
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US/60/236,033
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,032
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,028
; PRIOR FILING DATE: 2000-09-28
; NUMBER OF SEQ ID NOS: 583
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO: 273
; LENGTH: 167343
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-964-824A-273

Query Match 94.3%; Score 472.2; DB 10; Length 167343;
Best Local Similarity 96.4%; Pred. No. 2.9e-120;
Matches 483; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
QY 1 CTTTGGTGTATAGACATGAAGTCCCTTGCCCATGCCCTATGCTCCTGAATGCTATTGCGTG 60
|||||
Db 59330 CTTTGGTGTATAGACATGAAGTCCCTTGCCCATGCCCTATGCTCCTGAAGGTAATGCCCTA 59271
QY 61 GGTTCCTCTAGGATTTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTT 120
|||||
Db 59270 GGTTCCTCTAGGATTTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTT 59211
QY 121 AATTAATTTTGTATAGGCTAAGGATCCAGTTTCAAGCTTTTAACTTTTAACTTTTAACTTT 180
|||||
Db 59210 AATTAATTTTGTATAGGCTAAGGATCCAGTTTCAAGCTTTTAACTTTTAACTTTTAACTTT 59151
QY 181 GCCAGTTTCTCAGCACCATTTATTAATAGGGAATCCTTTCCCATGCTTTGTTTCT 240
|||||
Db 59150 GCCAGTTTCTCAGCACCATTTATTAATAGGGAATCCTTTCCCATGCTTTGTTTCT 59091
QY 241 CAGGTTTGTCAAAGATCAGATAGTTAGATATCGGGCATTTATTTCTGAGGCTCTGTTTC 300
|||||
Db 59090 CAGGTTTGTCAAAGATCAGATAGTTAGATATCGGGCATTTATTTCTGAGGCTCTGTTTC 59031
QY 301 TGTTCATTTGGTGTATATCTCTGTTTGGTACCAGTACCAGTGTGTTTGGTTACTGTAG 360
|||||
Db 59030 TGTTCATTTGGTGTATATCTCTGTTTGGTACCAGTACCAGTGTGTTTGGTTACTGTAG 58971
QY 361 CTTGTAGTGTAGTTGAAGTCAGGTACATGATGCCCTCCAGCTTTGTTCTTTGGGCTTA 420
|||||
Db 58970 CTTGTAGTGTAGTTGAAGTCAGGTACATGATGCCCTCCAGCTTTGTTCTTTGGGCTTA 58911
QY 421 GGATTGACTTGGCGATGGGCTCTTTTGGTTCACATGAACCTTTAAAGTAGTTTTT 480
|||||
Db 58910 GGATTGACTTGGCAATGGGCTCTTTTGGTTCACATGAACCTTTAAAGTAGTTTTT 58851
QY 481 CCAATTCTGTGAAGAAAGTAA 501
|||||
Db 58850 CCAATTCTGTGAAGAAAGTCA 58830

RESULT 8

US-10-003-806-6
; Sequence 6, Application US/10003806
; Patent No. US20020119929A1
; GENERAL INFORMATION:
; APPLICANT: Bishop, Colin E.
; APPLICANT: Agoulnik, Alexander I.
; APPLICANT: Zhu, Qichao
; TITLE OF INVENTION: CAN 1 AND ITS ROLE IN MAMMALIAN INFERTILITY
; FILE REFERENCE: P020660US1/10024824
; CURRENT APPLICATION NUMBER: US/10/003,806
; PRIOR FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 60/245,872
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 6
; LENGTH: 180557
; TYPE: DNA
; ORGANISM: Human
US-10-003-806-6

Query Match 93.6%; Score 469; DB 12; Length 180557;
Best Local Similarity 96.0%; Pred. No. 2.3e-119;
Matches 481; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY	1	CTTTTGGTCTTTAGACATGAAGTCTTGCCTCCCATGCTATGCTCCTGAATGGTATGCGCTG	60
Db	70311	CTTTTGGTCTTTAGACATGAAGTCTTGCCTCCCATGCTATGCTCCTGAATGGTATGCGCTA	70370
QY	61	GGTTTCTCTAGGATTTTATGGTTTAAAGTCTTAACATTTAAGTCTTTAAATCCCATCTTG	120
Db	70371	GGTTTCTCTAGGATTTTATGGTTTAAAGTCTTAACATTTAAGTCTTTAAATCCCATCTTG	70430
QY	121	AATTAATTTTGTATAAGTGTAAAGGAGGATCCAGTTTCAGCTTTACATATAGGGCTA	180
Db	70431	AATTAATTTTGTATAAGTGTAAAGGAGGATCCAGTTTCAGCTTTACATATAGGGCTA	70490
QY	181	GCAGTTTCTCAGACACCAATTTTAAATAGGAATCCCTTCCCATGCTTTGTTTCT	240
Db	70491	GCAGTTTCTCAGACACCAATTTTAAATAGGAATCCCTTCCCATGCTTTGTTTCT	70550
QY	241	CAGTTTGTCAAGATCAGATAGTGTAGATATGCGGCAATTTTCTGAGGCTCTGTTC	300
Db	70551	CAGTTTGTCAAGATCAGATAGTGTAGATATGCGGCTTATTTCTGAGGCTCTGTTC	70610
QY	301	TGTTCCATTTGATATCTCTGTTTGTACAGTACCATGCTTTGTTGTTTACTGTAG	360
Db	70611	TGTTCCATTTGATATCTCTGTTTGTACAGTACCATGCTTTGTTGTTTACTGTAG	70670
QY	361	CCTTGTAGTGTAGTTGAAGTCAAGTATGCGGCAATTTTCTGAGGCTCTGTTC	420
Db	70671	CCTTGTAGTGTAGTTGAAGTCAAGTATGCGGCAATTTTCTGAGGCTCTGTTC	70730
QY	421	GGATTGATTTGCAATGTGGCTCTTTTGGTTCACATGAACCTTAAAGTAGTTTTT	480
Db	70731	GGATTGATTTGCAATGTGGCTCTTTTGGTTCACATGAACCTTAAAGTAGTTTTT	70790
QY	481	CCAATTTCTGTGAAGAAAGTAA 501	
Db	70791	CCAATTTCTGTGAAGAAAGTCA 70811	

RESULT 9

US-10-003-806-9
; Sequence 9, Application US/10003806
; Patent No. US20020119929A1
; GENERAL INFORMATION:
; APPLICANT: Bishop, Colin E.
; APPLICANT: Agoulnik, Alexander I.
; APPLICANT: Zhu, Qichao

; TITLE OF INVENTION: CAN 1 AND ITS ROLE IN MAMMALIAN INFERTILITY
; FILE REFERENCE: P020660US1/10024824
; CURRENT APPLICATION NUMBER: US/10/003,806
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 60/245,872
; PRIOR FILING DATE: 2000-11-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 9
; LENGTH: 180557
; TYPE: DNA
; ORGANISM: Human
US-10-003-806-9

Query Match 93.6%; Score 469; DB 12; Length 180557;
Best Local Similarity 96.0%; Pred. No. 2.3e-119;
Matches 481; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY	1	CTTTTGGTCTTTAGACATGAAGTCTTGCCTCCCATGCTATGCTCCTGAATGGTATGCGCTG	60
Db	70311	CTTTTGGTCTTTAGACATGAAGTCTTGCCTCCCATGCTATGCTCCTGAATGGTATGCGCTA	70370
QY	61	GGTTTCTCTAGGATTTTATGGTTTAAAGTCTTAACATTTAAGTCTTTAAATCCCATCTTG	120
Db	70371	GGTTTCTCTAGGATTTTATGGTTTAAAGTCTTAACATTTAAGTCTTTAAATCCCATCTTG	70430
QY	121	AATTAATTTTGTATAAGTGTAAAGGAGGATCCAGTTTCAGCTTTACATATAGGGCTA	180
Db	70431	AATTAATTTTGTATAAGTGTAAAGGAGGATCCAGTTTCAGCTTTACATATAGGGCTA	70490
QY	181	GCAGTTTCTCAGACACCAATTTTAAATAGGAATCCCTTCCCATGCTTTGTTTCT	240
Db	70491	GCAGTTTCTCAGACACCAATTTTAAATAGGAATCCCTTCCCATGCTTTGTTTCT	70550
QY	241	CAGTTTGTCAAGATCAGATAGTGTAGATATGCGGCAATTTTCTGAGGCTCTGTTC	300
Db	70551	CAGTTTGTCAAGATCAGATAGTGTAGATATGCGGCTTATTTCTGAGGCTCTGTTC	70610
QY	301	TGTTCCATTTGATATCTCTGTTTGTACAGTACCATGCTTTGTTGTTTACTGTAG	360
Db	70611	TGTTCCATTTGATATCTCTGTTTGTACAGTACCATGCTTTGTTGTTTACTGTAG	70670
QY	361	CCTTGTAGTGTAGTTGAAGTCAAGTATGCGGCAATTTTCTGAGGCTCTGTTC	420
Db	70671	CCTTGTAGTGTAGTTGAAGTCAAGTATGCGGCAATTTTCTGAGGCTCTGTTC	70730
QY	421	GGATTGATTTGCAATGTGGCTCTTTTGGTTCACATGAACCTTAAAGTAGTTTTT	480
Db	70731	GGATTGATTTGCAATGTGGCTCTTTTGGTTCACATGAACCTTAAAGTAGTTTTT	70790
QY	481	CCAATTTCTGTGAAGAAAGTAA 501	
Db	70791	CCAATTTCTGTGAAGAAAGTCA 70811	

RESULT 10

US-09-764-891-8497/C
; Sequence 8497, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 8497
; LENGTH: 6684
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-8497

QY 61 GGTCTTCTCTAGGAGATTTATAGGTTTATAGGCTTAACATTTAAAGCTTTTAAATCCATCTTG 120
DB 13237 GGTCTTCTCTAGGAGATTTATAGGTTTATAGGCTTTAAAGCTTTTAAATCCATCTTG 13296
QY 121 AATTAATTTCTATAAGGTGTAAGGAAGGGATCCAGTTTTCAGCTTTCTACATAGGGCTA 180
DB 13297 AATTGATTTTCTATAAGGTGTAAGGAAGGGATCCAGTTTTCAGCTTTCTACATAGGGCTA 13356
QY 181 GCCAGTTTCTCAGCACCATTTTAAATAGGGAATCCCTTTCCCATTTGCTTTGTTTTCT 240
DB 13357 GCCAGTTTCCAGCACCATTTTAAATAGGGAATCCCTTTCCCATTTGCTTTGTTTTCT 13416
QY 241 CAGGTTTGCAAGATCAGATAGTTGTAGATATGCGGCAATATTTCTGAGGCTCTGTTC 300
DB 13417 CAGGTTTGCAAGATCAGATAGTTGTAGATATGCGGCAATATTTCTGAGGCTCTGTTC 13476
QY 301 TCTTCCATTTGTTGATATCTCTGTTTGGTACCACTACCATGTTGTTTGGTTACTGTAG 360
DB 13477 TCTTCCATTTGTTGATATCTCTGTTTGGTACCACTACCATGTTGTTTGGTTACTGTAG 13536
QY 361 CTTGTAGTGTAGTTTGAAGTCAGGTAGCATGCTCCAGCTTTGTTTCTTTTGGCTTA 420
DB 13537 CTTGTAGTGTAGTTTGAAGTCAGGTAGCATGCTCCAGCTTTGTTTCTTTTGGCTTA 13596
QY 421 GGATTGACTTGGCGATGCGGCTCTTTTGGTCCACATGAACCTTTAAAGTAGTTTTT 480
DB 13597 GGATTGACTTGGCGATGCGGCTCTTTTGGTCCACATGAACCTTTAAAGTAGTTTTT 13656
QY 481 CCAATTCCTGTGAAGAAAGTAA 501
DB 13657 CCAATTCCTGTGAAGAAAGTCA 13677

RESULT 13
US-09-728-721-63
; Sequence 63, Application US/09728721
; Patent No. US20020061845A1
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED PROTEIN FAMILY AND USES THERE
; CURRENT FILING DATE: 2000-12-01
; CURRENT APPLICATION NUMBER: US/09/728,721
; PRIOR FILING DATE: 1998-06-28
; PRIOR APPLICATION NUMBER: 09/340,620
; PRIOR FILING DATE: 1998-06-28
; PRIOR APPLICATION NUMBER: US/09/207,359
; PRIOR FILING DATE: 1998-12-08
; PRIOR APPLICATION NUMBER: US/09/099,041
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: US/09/019,942
; PRIOR FILING DATE: 1998-02-06
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 63
; LENGTH: 32042
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-728-721-63

Query Match 93.3%; Score 467.4; DB 10; Length 32042;
Best Local Similarity 95.8%; Pred. No. 2.8e-119;
Matches 480; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
QY 1 CTTTGGTCTTTTACACATGAAGTCTTGGCCCATGCTATGTCCTGAATGGTATGCGCTG 60
DB 13177 CTTTGGTCTTTTACACATGAAGTCTTGGCCCATGCTATGTCCTGAATGGTATGCGCTA 13236
QY 61 GGTCTTCTCTAGGAGATTTATAGGTTTATAGGCTTAACATTTAAAGCTTTTAAATCCATCTTG 120
DB 13237 GGTCTTCTCTAGGAGATTTATAGGTTTATAGGCTTTAAAGCTTTTAAATCCATCTTG 13296
QY 121 AATTAATTTTCTATAAGGTGTAAGGAAGGGATCCAGTTTTCAGCTTTCTACATAGGGCTA 180
DB 13297 AATTGATTTTCTATAAGGTGTAAGGAAGGGATCCAGTTTTCAGCTTTCTACATAGGGCTA 13356

DB 13297 AATTGATTTTCTATAAGGTGTAAGGAAGGGATCCAGTTTTCAGCTTTCTACATAGGGCTA 13356
QY 181 GCCAGTTTCTCAGCACCATTTTAAATAGGGAATCCCTTTCCCATTTGCTTTGTTTTCT 240
DB 13357 GCCAGTTTCCAGCACCATTTTAAATAGGGAATCCCTTTCCCATTTGCTTTGTTTTCT 13416
QY 241 CAGGTTTGCAAGATCAGATAGTTGTAGATATGCGGCAATATTTCTGAGGCTCTGTTC 300
DB 13417 CAGGTTTGCAAGATCAGATAGTTGTAGATATGCGGCAATATTTCTGAGGCTCTGTTC 13476
QY 301 TCTTCCATTTGTTGATATCTCTGTTTGGTACCACTACCATGTTGTTTGGTTACTGTAG 360
DB 13477 TCTTCCATTTGTTGATATCTCTGTTTGGTACCACTACCATGTTGTTTGGTTACTGTAG 13536
QY 361 CTTGTAGTGTAGTTTGAAGTCAGGTAGCATGCTCCAGCTTTGTTTCTTTTGGCTTA 420
DB 13537 CTTGTAGTGTAGTTTGAAGTCAGGTAGCATGCTCCAGCTTTGTTTCTTTTGGCTTA 13596
QY 421 GGATTGACTTGGCGATGCGGCTCTTTTGGTCCACATGAACCTTTAAAGTAGTTTTT 480
DB 13597 GGATTGACTTGGCGATGCGGCTCTTTTGGTCCACATGAACCTTTAAAGTAGTTTTT 13656
QY 481 CCAATTCCTGTGAAGAAAGTAA 501
DB 13657 CCAATTCCTGTGAAGAAAGTCA 13677

RESULT 14
US-10-006-883A-15
; Sequence 15, Application US/10006883A
; Publication No. US20030119767A1
; GENERAL INFORMATION:
; APPLICANT: Kenneth W. Dobie
; TITLE OF INVENTION: ANTISENSE MODULATION OF NOD1 EXPRESSION
; FILE REFERENCE: RTS-0337
; CURRENT APPLICATION NUMBER: US/10/006,883A
; CURRENT FILING DATE: 2001-12-05
; NUMBER OF SEQ ID NOS: 96
; SEQ ID NO 15
; LENGTH: 34001
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (425)...(524)
; OTHER INFORMATION: unknown
; FEATURE:
; NAME/KEY: intron:exon junction
; LOCATION: (10779)...(10780)
; OTHER INFORMATION: intron 1:exon 2
; FEATURE:
; NAME/KEY: intron:exon junction
; LOCATION: (9546)...(9547)
; OTHER INFORMATION: intron 7b:exon 8b
; FEATURE:
; NAME/KEY: exon
; LOCATION: (10878)...(10961)
; OTHER INFORMATION: exon 8
; FEATURE:
; NAME/KEY: intron:exon junction
; LOCATION: (21846)...(21847)
; OTHER INFORMATION: intron 10:exon 11
; FEATURE:
; NAME/KEY: exon:intron junction
; LOCATION: (21930)...(21931)
; OTHER INFORMATION: exon 11:intron 11
; FEATURE:
; NAME/KEY: intron
; LOCATION: (21931)...(24748)
; OTHER INFORMATION: intron 11
; FEATURE:
; NAME/KEY: exon:intron junction
; LOCATION: (29424)...(29425)

OTHER INFORMATION: exon 14:intron 14
FEATURE:
NAME/KEY: Intron
LOCATION: (29425)...(32217)
OTHER INFORMATION: Intron 14
FEATURE:
NAME/KEY: Intron:exon junction
LOCATION: (32217)...(32218)
OTHER INFORMATION: Intron 14:exon 15
FEATURE:
NAME/KEY: exon
LOCATION: (32218)...(33394)
OTHER INFORMATION: exon 15
US-10-067-514-15

Query Match 93.3%; Score 467.4; DB 9; Length 34001;
Best Local Similarity 95.8%; Pred. No. 2.9e-119;
Matches 480; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Qy 1 CTTTGGTGTATAGACATGAAGTCCTTGCCCATGCCCTATGCTCGAATGCTATGCGCTG 60
Db CTTTGGTGTATAGACATGAAGTCCTTGCCCATGCCCTATGCTCGAATGCTATGCGCTG 13758
Qy 61 GGTTCCTCTAGGAGTTTATGTTTATGCTCTAACATTTAAGTCTTTAATCCATCTTG 120
Db GGTTCCTCTAGGAGTTTATGTTTATGCTCTAACATTTAAGTCTTTAATCCATCTTG 13818
Qy 121 AATTAATTTTGTATAGGCTGAAGGAGGATCCAGTTTCAGCTTTCTACATAGGGCTA 180
Db AATTAATTTTGTATAGGCTGAAGGAGGATCCAGTTTCAGCTTTCTACATAGGGCTA 13878
Qy 181 GCCAGTTTCTCAGCACCAATTTAATAAGGGAATCCTTTCCCATGCTTTGTTTTCT 240
Db GCCAGTTTCTCAGCACCAATTTAATAAGGGAATCCTTTCCCATGCTTTGTTTTCT 13938
Qy 241 CAGTTTCTCAAGATCAGATAGTTGTAGATATCGGCATTTATTTCTGAGGGCTCTGTC 300
Db CAGTTTCTCAAGATCAGATAGTTGTAGATATCGGCATTTATTTCTGAGGGCTCTGTC 13998
Qy 301 TGTTCATTTGTTATCTCTGTTTGGTACCAGTACCATGTTGTTTGGTTACTGTAG 360
Db TGTTCATTTGTTATCTCTGTTTGGTACCAGTACCATGTTGTTTGGTTACTGTAG 14058
Qy 361 CCTTGTAGTGTAGTTGAAGTCAGTACGATGATGCCCTCCAGCTTTGTTCTTTGGCTTA 420
Db CCTTGTAGTGTAGTTGAAGTCAGTACGATGATGCCCTCCAGCTTTGTTCTTTGGCTTA 14118
Qy 421 GGATTGACTTGGCGATGCGGCTCTTTTGGTTCCACATGAACCTTTAAAGTAGTTTTT 480
Db GGATTGACTTGGCGATGCGGCTCTTTTGGTTCCACATGAACCTTTAAAGTAGTTTTT 14178
Qy 481 CCAATTCCTGGAAGAAAGTAA 501
Db CCAATTCCTGGAAGAAAGTCA 14199

RESULT 15
US-10-067-514-1/c
Sequence 1, Application US/10067514
Publication No. US20030054531A1
GENERAL INFORMATION:
APPLICANT: Jonsdottir, Solveig
APPLICANT: Jonsdottir, Sif
APPLICANT: Reynisdottir, Sigríður Th.
TITLE OF INVENTION: HUMAN STROKE GENE
FILE REFERENCE: 2345-2010-003
CURRENT APPLICATION NUMBER: US/10/067,514
PRIOR FILING DATE: 2002-02-04
PRIOR APPLICATION NUMBER: US 09/811/352
PRIOR FILING DATE: 2001-03-19
NUMBER OF SEQ ID NOS: 84
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1

LENGTH: 1691139
TYPE: DNA
ORGANISM: Human
US-10-067-514-1

Query Match 93.3%; Score 467.4; DB 9; Length 1691139;
Best Local Similarity 95.8%; Pred. No. 1.9e-118;
Matches 480; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Qy 1 CTTTGGTGTATAGACATGAAGTCCTTGCCCATGCCCTATGCTCGAATGCTATGCGCTG 60
Db CTTTGGTGTATAGACATGAAGTCCTTGCCCATGCCCTATGCTCGAATGCTATGCGCTA 59202
Qy 61 GGTTCCTCTAGGAGTTTATGTTTATGCTCTAACATTTAAGTCTTTAATCCATCTTG 120
Db GGTTCCTCTAGGAGTTTATGTTTATGCTCTAACATTTAAGTCTTTAATCCATCTTG 59142
Qy 121 AATTAATTTTGTATAGGCTGAAGGAGGATCCAGTTTCAGCTTTCTACATAGGGCTA 180
Db AATTAATTTTGTATAGGCTGAAGGAGGATCCAGTTTCAGCTTTCTACATAGGGCTA 59082
Qy 181 GCCAGTTTCTCAGCACCAATTTAATAAGGGAATCCTTTCCCATGCTTTGTTTTCT 240
Db GCCAGTTTCTCAGCACCAATTTAATAAGGGAATCCTTTCCCATGCTTTGTTTTCT 59022
Qy 241 CAGTTTGTCAAAGATCAGATAGTTGTAGATATCGGCATTTATTTCTGAGGGCTCTGTC 300
Db CAGTTTGTCAAAGATCAGATAGTTGTAGATATCGGCATTTATTTCTGAGGGCTCTGTC 58962
Qy 301 TGTTCATTTGTTATCTCTGTTTGGTACCAGTACCATGTTGTTTGGTTACTGTAG 360
Db TGTTCATTTGTTATCTCTGTTTGGTACCAGTACCATGTTGTTTGGTTACTGTAG 58902
Qy 361 CCTTGTAGTGTAGTTGAAGTCAGTACGATGATGCCCTCCAGCTTTGTTCTTTGGCTTA 420
Db CCTTGTAGTGTAGTTGAAGTCAGTACGATGATGCCCTCCAGCTTTGTTCTTTGGCTTA 58842
Qy 421 GGATTGACTTGGCGATGCGGCTCTTTTGGTTCCACATGAACCTTTAAAGTAGTTTTT 480
Db GGATTGACTTGGCGATGCGGCTCTTTTGGTTCCACATGAACCTTTAAAGTAGTTTTT 58782
Qy 481 CCAATTCCTGGAAGAAAGTAA 501
Db CCAATTCCTGGAAGAAAGTCA 58761

Search completed: July 4, 2003, 19:36:19
Job time : 118.075 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 15:16:37 ; Search time 35.0233 Seconds
(without alignments)
4386.938 Million cell updates/sec

Title: US-10-083-853B-2_COPY_29421_29921
Perfect score: 501
Sequence: 1 ctttgggttttagacatg.....caattctgtgaagaagtaa 501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Issued Patents_NA: *
1: /cgn2.6/ptodata/1/ina/5A_COMB.seq: *
2: /cgn2.6/ptodata/1/ina/5B_COMB.seq: *
3: /cgn2.6/ptodata/1/ina/6A_COMB.seq: *
4: /cgn2.6/ptodata/1/ina/6B_COMB.seq: *
5: /cgn2.6/ptodata/1/ina/PCTUS_COMB.seq: *
6: /cgn2.6/ptodata/1/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	467.4	93.3	32042	4	US-09-245-281-44
2	461.2	92.1	40328	3	US-08-742-185-102
3	452.4	90.3	18596	4	US-09-318-448-11
4	442.6	88.3	168575	4	US-09-426-290-1
5	440.4	87.9	10607	1	US-08-078-090-3
6	410.2	81.9	50000	4	US-09-146-053-4
7	319.2	63.7	10754	2	US-08-966-958-1
8	319.2	63.7	10754	2	US-09-215-817-1
9	319.2	63.7	10754	4	US-09-342-353-1
10	266.2	53.1	152331	3	US-09-128-155-16
11	266.2	53.1	176373	3	US-09-128-155-17
12	248.4	49.6	80246	4	US-09-078-294-4
13	248.4	49.6	80595	4	US-09-078-294-3
14	246.8	49.3	50000	4	US-09-146-053-3
15	246.2	49.1	38844	4	US-09-734-675-3
16	229.4	45.8	168575	4	US-09-426-290-1
17	187.2	37.4	22067	4	US-09-820-001-3
18	185	36.9	112132	4	US-09-741-150-3
19	184	36.7	98844	4	US-09-791-211-10
20	181	36.1	29604	3	US-08-781-891-207
21	174	34.7	624	4	US-09-385-982-209
22	165.2	33.0	1200	4	US-09-018-584A-37
23	161	32.1	10409	1	US-08-772-440-33
24	150.6	30.1	6367	1	US-08-470-299-1
25	150.6	30.1	6367	3	US-08-776-511-3
26	150.6	30.1	6926	1	US-08-470-299-2
27	144.8	28.9	55827	4	US-09-813-133A-3

c 28	133.2	26.6	3450	4	US-09-462-561B-10	Sequence 10, Appl
c 29	124.6	24.9	112132	4	US-09-741-150-3	Sequence 3, Appl
c 30	120.6	24.1	3430	4	US-09-462-561B-9	Sequence 9, Appl
c 31	119.8	23.9	4693	4	US-09-462-561B-8	Sequence 8, Appl
c 32	119.8	23.9	38564	4	US-09-734-673-3	Sequence 3, Appl
c 33	110.6	22.1	36159	4	US-09-749-588-3	Sequence 3, Appl
c 34	103.6	20.7	29604	3	US-08-781-891-207	Sequence 207, App
c 35	98	19.6	694	4	US-09-605-785-604	Sequence 604, App
c 36	98	19.6	694	4	US-09-605-785-665	Sequence 665, App
c 37	95.8	19.1	90050	4	US-09-245-041-5	Sequence 5, Appl
c 38	95.6	19.1	176373	3	US-09-128-155-17	Sequence 17, Appl
c 39	95.2	19.0	43793	3	US-08-742-185-101	Sequence 101, App
c 40	90.4	18.0	4344	4	US-09-462-561B-11	Sequence 11, Appl
c 41	90.2	18.0	705	4	US-09-605-785-666	Sequence 666, App
c 42	89.4	17.8	11811	4	US-09-078-294-7	Sequence 7, Appl
c 43	88.6	17.7	246240	2	US-08-724-394A-20	Sequence 20, Appl
c 44	88.6	17.7	246240	2	US-08-724-394A-21	Sequence 21, Appl
c 45	88.6	17.7	246240	2	US-08-724-394A-22	Sequence 22, Appl

ALIGNMENTS

RESULT 1
US-09-245-281-44
; Sequence 44, Application US/09245281
; Patent No. 6369196
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED PROTEIN FAMILY
; FILE REFERENCE: 07334/118001
; CURRENT APPLICATION NUMBER: US/09/245,281
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: US 09/207,359
; EARLIER FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: US 09/099,041
; EARLIER FILING DATE: 1998-06-17
; EARLIER APPLICATION NUMBER: US 09/019,942
; EARLIER FILING DATE: 1998-02-06
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 32042
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-245-281-44

Query Match 93.3%; Score 467.4; DB 4; Length 32042;
Best Local Similarity 95.8%; Pred.No. 6e-129;
Matches 480; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 CTTTGGTGTTTAGACATGAAGTCTCTGCCCATGCTATGCTCCTGAATGGTAAATGCTG 60
DB 13177 CTTTGGTGTTTGGACATGAAGTCTCTGCCCATGCTATGCTCCTGAATGGTAAATGCTG 13236
QY 61 GGTTCCTCTAGGGATTTTATGGTTTAACTTAAGTCTTAACTTAAGTCTTAACTTAAGTCTT 120
DB 13237 GGTTCCTCTAGGGATTTTATGGTTTAACTTAAGTCTTAACTTAAGTCTTAACTTAAGTCTT 13296
QY 121 AATTAATTTTCTATAGGTGTAAGGAAGGATCCAGTTTCAGCTTTTACATAGGCTA 180
DB 13297 AATTGATTTTGTATAAGGTGTAAGGAAGGATCCAGTTTCAGCTTTTACATAGGCTA 13356
QY 181 GCAGTTTCTCAGACACCATTTTAAATAGGAATCCCTTCCCATTCCTTCCCATTCCTTCCCATTCCTT 240
DB 13357 GCAGTTTCTCAGACACCATTTTAAATAGGAATCCCTTCCCATTCCTTCCCATTCCTTCCCATTCCTT 13416
QY 241 CAGGTTTGTCAAGATCAGATAGTTCTAGATATGCGGCATTTATTTCTGAGGCTCTGTTTC 300
DB 13417 CAGGTTTGTCAAGATCAGATAGTTCTAGATATGCGGCATTTATTTCTGAGGCTCTGTTTC 13476
QY 301 TGTTCATTCGTTGATATCTCTGTTTGTACCAGTACCATGTTGTTTGTGTTTGTGTTTGTGTTTGTGTTGTTAG 360

Db 13477 TGTTCATGTATCTCTGTTTGGTACCATGCTGTTTGGTTACTGTAG 13536
Qy 361 CCTTGTAGTGTAGTTGAAGTCAGTCATGATGCTCCAGCTTTGTCTTTTGGCTTA 420
Db 13537 CCTGTAGTATAGTTGAAGTCAGTCATGATGCTCCAGCTTTGTCTTTTGGCTTA 13596
Qy 421 GGATTGACTTGGCGATGCGGCTCTTTTTTGGTTCCACATGAACCTTTAAAGTAGTTTTT 480
Db 13597 GGATTGACTTGGCGATGCGGCTCTTTTTTGGTTCCATATGAACCTTTAAAGTAGTTTTT 13656
Qy 481 CCAATTCGTGAAGAAAGTAA 501
Db 13657 CCAATTCGTGAAGAAAGTCA 13677

RESULT 2

US-08-742-185-102
; Sequence 102, Application US/08742185
; Patent No. 6020476
; GENERAL INFORMATION:
; APPLICANT: Page, David C.
; APPLICANT: Reijo, Renee
; APPLICANT: Saxena, Richa
; APPLICANT: Hawkins, Trevor
; APPLICANT: Reeve, Mary Pat
; TITLE OF INVENTION: DAZ: A GENE FAMILY ASSOCIATED WITH AZOOSPERMIA
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: US
; ZIP: 02173

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/742,185
; FILING DATE: 30-OCT-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/690,734
; FILING DATE: 31-JUL-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/310,429
; FILING DATE: 22-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: WHI94-07A2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 102:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 40328 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-742-185-102

Query Match 92.1%; Score 461.2; DB 3; Length 40328;
Best Local Similarity 95.4%; Pred. No. 4.6e-127;
Matches 475; Conservative 0; Mismatches 23; Indels 0; Gaps 0;
Qy 1 CTTTTGGTGTATAGACATGAAGTCCTTGGCCATGCTATGCTCTGAATGGTATTGCGTG 60
Db 17396 CTTTTGGTGTATAGACATGAAGTCCTTGGCCATGCTATGCTCTGAATGGTATTGCGTG 17455

Qy 61 GGTTCCTCTCCTAGGATTTTATGTTTGTAGCTTAACATTTAAGTCTTTAATCCATCTTG 120
Db 17456 GGTTCCTCTCCTAGGATTTTATGTTTGTAGCTTAACATTTAAGTCTTTAATCCATCTTG 17515
Qy 121 AATTAATTTTGTATAGGTGTAAAGGAGGATCCAGTTTCAGCTTTCTACATAGGGCTA 180
Db 17516 AATTAATTTTGTATAGGTGTAAAGGAGGATCCAGTTTCAGCTTTCTACATAGGGCTA 17575
Qy 181 GCCAGTTTCTCAGCACCATTTATTAATAGGGAATCCTTTCCCATGCTTTGTTTCT 240
Db 17576 GCCAGTTTCTCAGCACCATTTATTAATAGGGAATCCTTTCCCATGCTTTGTTTCT 17635
Qy 241 CAGTTTGTCAAAGATCAGATAGTTGTAGATATCGGCATTTATTTCTGAGGGCTCTGTTC 300
Db 17636 CAGTTTGTCAAAGATCAGATAGTTGTAGATATCGGCATTTATTTCTGAGGGCTCTGTTC 17695
Qy 301 TGTTCATTTGTTGATATCTCTGTTTGGTACCAGTACCAGTGTGTTTGGTTACTGTAG 360
Db 17696 TGTTCATTTGTTGATATCTCTGTTTGGTACCAGTACCAGTGTGTTTGGTTACTGTAG 17755
Qy 361 CCTGTAGTGTAGTTTGAAGTCAGTAGCATGATGCTCCAGCTTTGTTCTTTTGGCTTA 420
Db 17756 CCTGTAGTGTAGTTTGAAGTCAGTAGGTGTGTCCTCCAGCTTTGTTCTTTTGGCTTA 17815
Qy 421 GGATTGACTTGGCGATGCGGCTCTTTTTTGGTTCCACATGAACCTTTAAAGTAGTTTTT 480
Db 17816 GGATTGACTTGGCGATGCGGCTCTTTTTTGGTTCCATGTGAACCTTTAAAGTAGTTTTT 17875
Qy 481 CCAATTCGTGAAGAAAG 498
Db 17876 CCAATTCGTGAAGAAAG 17893

RESULT 3

US-09-318-448-11
; Sequence 11, Application US/09318448
; Patent No. 6210950
; GENERAL INFORMATION:
; APPLICANT: Johnson, William G.
; APPLICANT: Stenroos, Edward S.
; TITLE OF INVENTION: METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
; FILE REFERENCE: 601-1-057
; CURRENT APPLICATION NUMBER: US/09/318,448
; CURRENT FILING DATE: 1999-05-25
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 18596
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-318-448-11

Query Match 90.3%; Score 452.4; DB 4; Length 18596;
Best Local Similarity 94.7%; Pred. No. 1.4e-124;
Matches 468; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
Qy 8 TGTTCATGACATGAAGTCCTTGGCCATGCTCTGCTGATGCTGATGCTGCTGGGTTTC 67
Db 6806 TGCATAGGCATGAAGTCCTTGGCCATGCTCTGCTGATGCTGATGCTGCTGGGTTTC 6865
Qy 68 TTCTAGGGATTTTATGTTTGTAGTCTAAACATTTAAGTCTTTAATCCATCTTGAATTAAT 127
Db 6866 TTCTAGGGATTTTATGTTTGTAGTCTAAACATTTAAGTCTTTAATCCATCTTGAATTAAT 6925
Qy 128 TTTTGTATAAGGTGTAAAGGAGGATCCAGTTTCCAGCTTTCTACATAGGGGTACCCAGTT 187
Db 6926 TTTTGTATAAGGTGTAAAGGAGGATCCAGTTTCCAGCTTTCTACATAGGGGTACCCAGTT 6985
Qy 188 TTCTCAGCACCATTTTAAATAGGGAATCCTTTCCCATGCTGTTTCTCAGGTTT 247
Db 6986 TTCCAGCACCATTTTAAATAGGGAATCCTTTCCCATGCTGTTTCTCAGGTTT 7045

```
QY 248 GTCAAGATCAGATAGTGTAGATATGCGCATATTTCTGAGGCTCTCTCTGTGTCCA 307
Db 7046 GTCAAGATCAGATAGTGTAGATATGCGCGTATTTCTGAGGCTCTCTCTGTGTCCA 7105
QY 308 TTGGTTCATATCTCTGTTTGGTACCAAGTACCAATGTTGTTGGTACTAGCTTGTA 367
Db 7106 TTGATCTATGCTGTGTTTGGTACCAAGTACCAATGTTGTTGGTACTAGCTTGTA 7165
QY 368 GTGTAGTGTCAAGTACAGTACGATGATGCGCTCCAGCTTTCTTTGGCTTAGGATTGA 427
Db 7166 GTATAGTGTCAAGTACAGTACGATGATGCGCTCCAGCTTTCTTTGGCTTAGGATTGA 7225
QY 428 CTTGGGATGCTGGCTCTTTTGGTGTCCACATGAACCTTTAAAGTAGTTTTTCCCAATTC 487
Db 7226 CTTGGGATGCTGGCTCTTTTGGTGTCCACATGAACCTTTAAAGTAGTTTTTCCCAATTC 7285
QY 488 TGTGAAGAAAGTAA 501
Db 7286 TGTGAAGAAAGTAA 7299

RESULT 4
US-09-426-290-1/c
; Sequence 1, Application US/09426290
; Patent No. 6410712
; GENERAL INFORMATION:
; APPLICANT: Berjilind Ran Olafsdottir
; APPLICANT: Jeffrey Gulcher
; TITLE OF INVENTION: HUMAN NARCOLEPSY GENE
; FILE REFERENCE: 2345.2001-000
; CURRENT APPLICATION NUMBER: US/09/426.290
; CURRENT FILING DATE: 1999-10-25
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 168575
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (21181)...(21403)
; NAME/KEY: CDS
; LOCATION: (95252)...(95430)
; NAME/KEY: CDS
; LOCATION: (101753)...(101996)
; NAME/KEY: CDS
; LOCATION: (110324)...(110439)
; NAME/KEY: CDS
; LOCATION: (124058)...(124278)
; NAME/KEY: CDS
; LOCATION: (127009)...(127130)
; NAME/KEY: CDS
; LOCATION: (128910)...(129139)
; US-09-426-290-1

Query Match 88.3%; Score 442.6; DB 4; Length 168575;
Best Local Similarity 94.08; Pred. No. 2.7e-121;
Matches 471; Conservative 0; Mismatches 29; Indels 1; Gaps 1;

QY 1 CTTTGGTGTGTTTAGACATGAAGTCCCTGTCGATGCTCTGCTGAATGCTATGCTG 60
Db 159839 CTTTGGTGTGTTTAGACATGAAGTCCCTGTCGATGCTCTGCTGAATGCTATG 159780
QY 61 GGTTCCTCTAGGATTTTATGGTTTAGCTTAACATTTAGCTTTTAATCCATCTTG 120
Db 159779 GGTTCCTCTAGGATTTTATGGTTTAGCTTAACATTTAGCTTTTAATCCATCTTG 159721
QY 121 AATTAATTTTGTATAGGTGTAAGGAAGGATCCAGTTTCAGCTTTCTACATAGGCTA 180
Db 159720 AGTTAATTTTGTATAGGTGTAAGGAAGGATCCAGTTTCAGCTTTCTACATAGGCTA 159661
QY 181 GCCAGTTTCTCAGCACCACTTTTAAATAGGAATCCTTTCCCAATGCTGTTTCT 240
```

```
Db 159660 ACCAGTTTCCAGCACCATTTATTAATAGGAATCCTTTCCCATTTGCTTTTCT 159601
QY 241 CAGTTTTGTCAAGATCAGATAGTGTAGATATGCGCATATTTCTGAGGCTCTGTTTC 300
Db 159600 CAGTTTTGTCAAGATCAGATAGTGTAGATATGCGCATATTTCTGAGGCTCTGTTTC 159541
QY 301 TGTTCATTTGTTGATATCTCTGTTTGGTACCAAGTACCAATGTTGTTTGGTACTGTAG 360
Db 159540 TGTTCATTTGTTGATATCTCTGTTTGGTACCAAGTACCAATGTTGTTTGGTACTGTAG 159481
QY 361 CTTTGTAGTGTGTTGAAAGTACAGTACGATGCTCCAGCTTTCTTTTGGCTTA 420
Db 159480 CTTTGTAGTGTGTTGAAAGTACAGTACGATGCTCCAGCTTTCTTTTGGCTTA 159421
QY 421 GGATTTGCTGGGATGCTGGGCTCTTTTGGTGTCCACATGAACCTTTAAAGTACTTTTT 480
Db 159420 GGATTTGCTGGGATGCTGGGCTCTTTTGGTGTCCACATGAACCTTTAAAGTACTTTTT 159361
QY 481 CCAATTTCTGTGAAGAAAGTAA 501
Db 159360 CCAATTTCTGTGAAGAAAGTAA 159340

RESULT 5
US-08-078-090-3
; Sequence 3, Application US/08078090
; Patent No. 5739407
; GENERAL INFORMATION:
; APPLICANT: BERGSTROEM, SVEN
; APPLICANT: HERNELL, OLLE
; APPLICANT: LOENNERDAL, BO
; APPLICANT: HJALMARSSON, KARIN
; APPLICANT: HANSSON, LENNART
; APPLICANT: TOERNELL, JAN
; APPLICANT: STROEMQUIST, MATS
; TITLE OF INVENTION: HUMAN BETA-CASEIN PROCESS FOR PRODUCING
; TITLE OF INVENTION: IT AND USE THEREOF
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 SEVENTH STREET, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/078,090
; FILING DATE: 19930618
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/DK92/00236
; FILING DATE: 19-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/DK91/00233
; FILING DATE: 19-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: COOPER, IVER P.
; REGISTRATION NUMBER: 28, 005
; REFERENCE/DOCKET NUMBER: BERGSTROM2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)628-5197
; TELEFAX: (202)737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10607 base pairs
; TYPE: nucleic acid
```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(4804..4854, 5720..5746, 6726..6746, 6845
; LOCATION: ..6886, 7991..8521, 9440..9445)
US-08-078-090-3

Query Match      87.9%; Score 440.4; DB 1; Length 10607;
Best Local Similarity 93.6%; Pred. No. 3.9e-121;
Matches 470; Conservative 0; Mismatches 31; Indels 1; Gaps 1;

QY 1 CTTTGGTGTAGACATGAAGTCTTGGCCATGCTATGCTCCTGAAGTATGCCTG 60
Db CTTTGGTGTAGACATGAAGTCTTGGCCATGCTATGCTCCTGAAGTATGCCTG 60
QY 61 GGTTCCTCTAGGATTTATAGTCTTAACATTTAAGTCTTTAATCCATCTTG 120
Db GGTTCCTCTAGGATTTATAGTCTTAACATTTAAGTCTTTAATCCATCTTG 120
QY 121 AATTAATTTTGTATAGGTGTAAGGAAGGATCCAGTTTCAGCTTCTTACATAGGCTA 180
Db AATTAATTTTGTATAGGTGTAAGGAAGGATCCAGTTTCAGCTTCTTACATAGGCTA 180
QY 27160 AATTAATTTTGTATAGGTGTAAGGAAGGATCCAGTTTCAGCTTCTTACATAGGCTA 27219
Db AATTAATTTTGTATAGGTGTAAGGAAGGATCCAGTTTCAGCTTCTTACATAGGCTA 27219
QY 181 GCCAGTTTCTCAGCAGCACCATTATTAATAGGAATCCTTTCCCATTCCTGTTTTTCT 240
Db GCCAGTTTCTCAGCAGCACCATTATTAATAGGAATCCTTTCCCATTCCTGTTTTTCT 240
QY 27220 GCCGTTTTCAGCAGCACCATTATTAATAGGAATCCTTTCCCATTCCTGTTTTTGT 27279
Db GCCGTTTTCAGCAGCACCATTATTAATAGGAATCCTTTCCCATTCCTGTTTTTGT 27279
QY 241 CAGGTTTGTCAAAGATCAGATAGTTGTAGATATCGGCATATTTCTGAGGGCTCTGTTTC 300
Db CAGGTTTGTCAAAGATCAGATAGTTGTAGATATCGGCATATTTCTGAGGGCTCTGTTTC 300
QY 27280 CAGGTTTGTCAAAGATCAGATAGTTGTAGATATCGGCATATTTCTGAGGGCTCTGTTTC 27339
Db CAGGTTTGTCAAAGATCAGATAGTTGTAGATATCGGCATATTTCTGAGGGCTCTGTTTC 27339
QY 301 TGTTCATTTGTTGATATCTCTGTTTTGTTGGTACCAGTACCATGTTGTTGGTACTGTAG 360
Db TGTTCATTTGTTGATATCTCTGTTTTGTTGGTACCAGTACCATGTTGTTGGTACTGTAG 360
QY 27340 TGTTCATTTGTTGATATCTCTGTTTTGTTGGTACCAGTACCATGTTGTTGGTACTGTAG 27399
Db TGTTCATTTGTTGATATCTCTGTTTTGTTGGTACCAGTACCATGTTGTTGGTACTGTAG 27399
QY 361 CCTGTAGTGTAGTCTTGAAGTCCAGTATGATGCTCCAGCTTTGTTCTTTTGGCTTA 420
Db CCTGTAGTGTAGTCTTGAAGTCCAGTATGATGCTCCAGCTTTGTTCTTTTGGCTTA 420
QY 421 GGATTCACATTTGGCGATGTTGGGCTCTTTTGGTTC 455
Db GGATTCACATTTGGCGATGTTGGGCTCTTTTGGTTC 455

RESULT 7
US-08-966-958-1/c
; Sequence 1, Application US/08966958
; Patent No. 5928908
; GENERAL INFORMATION:
; APPLICANT: Dunn, John
; APPLICANT: Randes, Matthew
; TITLE OF INVENTION: METHODS FOR INTRODUCING UNIDIRECTIONAL
; TITLE OF INVENTION: DELETIONS
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brookhaven National Laboratory
; STREET: P.O. Box 5000
; CITY: Upton
; STATE: New York
; COUNTRY: US
; ZIP: 11973
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/966,958
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bogosian, Margaret
; REGISTRATION NUMBER: 25,324
; REFERENCE/DOCKET NUMBER: AUI97-14
; TELEPHONE: (516) 344-3341
; TELEFAX: (516) 344-3729
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10754 base pairs
; TYPE: nucleic acid

Query Match      81.9%; Score 410.2; DB 4; Length 50000;
Best Local Similarity 93.6%; Pred. No. 3.9e-121;
Matches 470; Conservative 0; Mismatches 31; Indels 1; Gaps 1;

QY 1 CTTTGGTGTAGACATGAAGTCTTGGCCATGCTATGCTCCTGAAGTATGCCTG 60
Db CTTTGGTGTAGACATGAAGTCTTGGCCATGCTATGCTCCTGAAGTATGCCTG 60
QY 61 GGTTCCTCTAGGATTTATAGTCTTAACATTTAAGTCTTTAATCCATCTTG 120
Db GGTTCCTCTAGGATTTATAGTCTTAACATTTAAGTCTTTAATCCATCTTG 120
QY 121 AATTAATTTTGTATAGGTGTAAGGAAGGATCCAGTTTCAGCTTCTTACATAGGCTA 180
Db AATTAATTTTGTATAGGTGTAAGGAAGGATCCAGTTTCAGCTTCTTACATAGGCTA 180
QY 301 TGTTCATTTGTTGATATCTCTGTTTTGTTGGTACC - AGTACCATGTTGTTTGGTACTGTA 359
Db TGTTCATTTGTTGATATCTCTGTTTTGTTGGTACC - AGTACCATGTTGTTTGGTACTGTA 359
QY 361 CCTGTAGTGTAGTCTTGAAGTCCAGTATGATGCTCCAGCTTTGTTCTTTTGGCTTA 420
Db CCTGTAGTGTAGTCTTGAAGTCCAGTATGATGCTCCAGCTTTGTTCTTTTGGCTTA 420
QY 421 GGATTCACATTTGGCGATGTTGGGCTCTTTTGGTTC 455
Db GGATTCACATTTGGCGATGTTGGGCTCTTTTGGTTC 455

RESULT 6
US-09-146-053-4
; Sequence 4, Application US/09146053A
; Patent No. 6399349
; GENERAL INFORMATION:
; APPLICANT: Ryan, James W.
; APPLICANT: Sprinkle, Terry Joe Curtis
; APPLICANT: Venema, Richard C.
; TITLE OF INVENTION: Human Amino peptidase P Gene
; FILE REFERENCE: MCG103
; CURRENT APPLICATION NUMBER: US/09/146,053A
; CURRENT FILING DATE: 1998-09-02
; EARLIER APPLICATION NUMBER: 60/057,854
; EARLIER FILING DATE: 1997-09-02
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 50000
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-146-053-4
```


STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-966-958-1

Query Match 63.7%; Score 319.2; DB 2; Length 10754;
Best Local Similarity 94.8%; Pred. No. 3.8e-85;
Matches 330; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 154 CCAGTTTCAGCTTTCTACATAGGCGTAGCCAGTTTCTAGACACCATTTATTAATAGGG 213
DB 10754 CCAGTTTCAGCTTTCTATATATAGCGTAGCCAGTTTCCAGACACCATTTATTAATAGGG 10695
QY 214 AATCCTTCCCATCTGTTTCTCAGGTTTGCAAGATCAGATAGTTAGATAT 273
DB 10694 AATCCTTCCCATCTGTTTCTCAGGTTTGCAAGATCAGATAGTTAGATAT 10635
QY 274 GCGGCATTAATTTCTGAGGCGCTGTTCTGTCATTGGTTGATATCTGTTTGTACC 333
DB 10634 GTGGCGTTAATTTCTGAGGCGCTGTTCTGTCATTGGTTAATCTGTTTGTGCACC 10575
QY 334 AGTACCATGTTGTTTGTACTAGTACCTTGTAGTGTAGTTGAAGTCAGTACATGA 393
DB 10574 AATACCATGCTGTTTGTACTAGTACCTTGTAGTGTAGTTGAAGTCAGTACATGA 10515
QY 394 TGCCCTCAGCTTGTCTTTGGCTTAGGATGACTTGGCGATGTGGCTCTTTTGGT 453
DB 10514 TGCCCTCAGCTTGTCTTTGGCTTAGGATGACTTGGCGATGTGGCGCTCTTTTGGT 10455
QY 454 TCACATGAACCTTTAAAGTAGTTTTCATTCATTCGTGAAGAAATGA 501
DB 10454 TCATATGAACCTTTAAAGTAGTTTTCATTCATTCGTGAAGAAATGA 10407

RESULT 8
US-09-215-817-1/c

; Sequence 1, Application US/09215817
; Patent No. 5968786
; GENERAL INFORMATION:

; APPLICANT: Dunn, John
; TITLE OF INVENTION: METHODS FOR INTRODUCING UNIDIRECTIONAL
; TITLE OF INVENTION: DELETIONS
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brookhaven National Laboratory
; STREET: P.O. Box 5000
; CITY: Upton
; STATE: New York
; COUNTRY: US
; ZIP: 11973

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/215,817
; FILING DATE:

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/966,958
; FILING DATE:

; ATTORNEY/AGENT INFORMATION:
; NAME: Bogosian, Margaret
; REGISTRATION NUMBER: 25,324
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 344-3341
; TELEFAX: (516) 344-33729
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10754 base pairs

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-215-817-1

Query Match 63.7%; Score 319.2; DB 2; Length 10754;
Best Local Similarity 94.8%; Pred. No. 3.8e-85;
Matches 330; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 154 CCAGTTTCAGCTTTCTACATAGGCGTAGCCAGTTTCTAGACACCATTTATTAATAGGG 213
DB 10754 CCAGTTTCAGCTTTCTATATATAGCGTAGCCAGTTTCCAGACACCATTTATTAATAGGG 10695
QY 214 AATCCTTCCCATCTGTTTCTCAGGTTTGCAAGATCAGATAGTTAGATAT 273
DB 10694 AATCCTTCCCATCTGTTTCTCAGGTTTGCAAGATCAGATAGTTAGATAT 10635
QY 274 GCGGCATTAATTTCTGAGGCGCTGTTCTGTCATTGGTTGATATCTGTTTGTACC 333
DB 10634 GTGGCGTTAATTTCTGAGGCGCTGTTCTGTCATTGGTTAATCTGTTTGTGCACC 10575
QY 334 AGTACCATGTTGTTTGTACTAGTACCTTGTAGTGTAGTTGAAGTCAGTACATGA 393
DB 10574 AATACCATGCTGTTTGTACTAGTACCTTGTAGTGTAGTTGAAGTCAGTACATGA 10515
QY 394 TGCCCTCAGCTTGTCTTTGGCTTAGGATGACTTGGCGATGTGGCGCTCTTTTGGT 453
DB 10514 TGCCCTCAGCTTGTCTTTGGCTTAGGATGACTTGGCGATGTGGCGCTCTTTTGGT 10455
QY 454 TCACATGAACCTTTAAAGTAGTTTTCATTCATTCGTGAAGAAATGA 501
DB 10454 TCATATGAACCTTTAAAGTAGTTTTCATTCATTCGTGAAGAAATGA 10407

RESULT 9

US-09-342-353-1/c
; Sequence 1, Application US/09342353
; Patent No. 6248569
; GENERAL INFORMATION:

; APPLICANT: Dunn, John
; TITLE OF INVENTION: METHOD FOR INTRODUCING UNIDIRECTIONAL NESTED DELETIONS
; FILE REFERENCE: CIP OF U.S. Application 08/966,958
; CURRENT APPLICATION NUMBER: US/09/342,353
; CURRENT FILING DATE: 1999-06-29
; EARLIER APPLICATION NUMBER: 08/966,958
; EARLIER FILING DATE: 1997-11-10
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 10754
; TYPE: DNA
; ORGANISM: Homo sapiens

US-09-342-353-1

Query Match 63.7%; Score 319.2; DB 4; Length 10754;
Best Local Similarity 94.8%; Pred. No. 3.8e-85;
Matches 330; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 154 CCAGTTTCAGCTTTCTACATAGGCGTAGCCAGTTTCTAGACACCATTTATTAATAGGG 213
DB 10754 CCAGTTTCAGCTTTCTATATATAGCGTAGCCAGTTTCCAGACACCATTTATTAATAGGG 10695
QY 214 AATCCTTCCCATCTGTTTCTCAGGTTTGCAAGATCAGATAGTTAGATAT 273
DB 10694 AATCCTTCCCATCTGTTTCTCAGGTTTGCAAGATCAGATAGTTAGATAT 10635
QY 274 GCGGCATTAATTTCTGAGGCGCTGTTCTGTCATTGGTTGATATCTGTTTGTACC 333
DB 10634 GTGGCGTTAATTTCTGAGGCGCTGTTCTGTCATTGGTTAATCTGTTTGTGCACC 10575
QY 334 AGTACCATGTTGTTTGTACTAGTACCTTGTAGTGTAGTTGAAGTCAGTACATGA 393
DB 10407 TCATATGAACCTTTAAAGTAGTTTTCATTCATTCGTGAAGAAATGA 10407

Db 10574 AATACATGCTGTTTGGTACTAGTAGGCTTGTAGTGTGAAGTCAGTACTGTCA 10515
Qy 394 TGGCTCCAGCTTTGTTCTTTGGCTTAGGATTGACCTGGCGCTCTTTTGGT 453
Db 10514 TGGCTCCAGCTTTGTTCTTTGGCTTAGGATTGACCTGGCGCTCTTTTGGT 10455
Qy 454 TCCCATGACCTTTAAAGTACTTTTCCCAATTCGTGTAAGAAATAA 501
Db 10454 TCCCATGACCTTTAAAGTACTTTTCCCAATTCGTGTAAGAAATAA 10407

RESULT 10

US-09-128-155-16/C
; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; EARLIER FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 152331
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: (1)...(152331)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

Query Match 53.1%; Score 266.2; DB 3; Length 152331;
Best Local Similarity 88.5%; Pred. No. 6.2e-69;
Matches 300; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

Qy 99 TTTAAGTCTTAATCCATCTTGAATTAATTTTGTATAGGTGTAAGAGGATCCAGT 158
Db 38182 TTTAATCTTAATCCATCTTGAATTAATTTTGTATAGGTGTAAGAGGATCCAGT 38123
Qy 159 TTCAGCTTTACATAGGCGTACGACGTTTCTCAGCAATTTTAAATAGGAA-TC 217
Db 38122 TTCAGCTTTACATAGGCGTACGACGTTTCTCAGCAATTTTAAATAGGAA-TC 38063
Qy 218 CTTTCCCATGCTTTTCTCAGGTTTGTCAAGATCAGATAGTGTAGATATGCGG 277
Db 38062 TTTTCCCATGCTTTTCTCAGGTTTGTCAAGATCAGATAGTGTAGATATGCGG 38003
Qy 278 CATATTTCTGAGGCGTCTGTTCTGTTCCATTTGTTGATCTCTGTTTGGTACAGTA 337
Db 38002 TGTATATCTGAGGCGTCTGTTCTGTTTGGTCTATATCTGTTTGGTACAGTA 37943
Qy 338 CCATGTTGTTTGGTACTGTAGCTTGTAGTGTGAAGTGAAGTACGATGATGCC 397
Db 37942 CCATGTTGTTTGGTACTGTAGCTTGTAGTGTGAAGTGAAGTACGATGATGCC 37883
Qy 398 TCCAGCTTTGTTCTTTGGCTTAGGATGACTTGGCAT 436
Db 37882 TCCAGCTTTGTTCTTTGGCTTAGGATGACTTGGCAT 37844

RESULT 11
US-09-128-155-17/C
; Sequence 17, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang

; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; EARLIER FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 176373
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: (1)...(176373)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-17

Query Match 53.1%; Score 266.2; DB 3; Length 176373;
Best Local Similarity 88.5%; Pred. No. 6.6e-69;
Matches 300; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

Qy 99 TTTAAGTCTTAATCCATCTTGAATTAATTTTGTATAGGTGTAAGAGGATCCAGT 158
Db 42560 TTTAATCTTAATCCATCTTGAATTAATTTTGTATAGGTGTAAGAGGATCCAGT 42501
Qy 159 TTCAGCTTTACATAGGCGTACGACGTTTCTCAGCAATTTTAAATAGGAA-TC 217
Db 42500 TTCAGCTTTACATAGGCGTACGACGTTTCTCAGCAATTTTAAATAGGAA-TC 42441
Qy 218 CTTTCCCATGCTTTTCTCAGGTTTGTCAAGATCAGATAGTGTAGATATGCGG 277
Db 42440 TTTTCCCATGCTTTTCTCAGGTTTGTCAAGATCAGATAGTGTAGATATGCGG 42381
Qy 278 CATATTTCTGAGGCGTCTGTTCTGTTCCATTTGTTGATCTCTGTTTGGTACAGTA 337
Db 42380 TGTATATCTGAGGCGTCTGTTCTGTTTGGTCTATATCTGTTTGGTACAGTA 42321
Qy 338 CCATGTTGTTTGGTACTGTAGCTTGTAGTGTGAAGTGAAGTACGATGATGCC 397
Db 42320 CCATGTTGTTTGGTACTGTAGCTTGTAGTGTGAAGTGAAGTACGATGATGCC 42261
Qy 398 TCCAGCTTTGTTCTTTGGCTTAGGATGACTTGGCAT 436
Db 42260 TCCAGCTTTGTTCTTTGGCTTAGGATGACTTGGCAT 42222

RESULT 12
US-09-078-294-4
; Sequence 4, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sarr, Desiree
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078,294
; EARLIER FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 80246
; TYPE: DNA
; ORGANISM: Nucleotide sequence of NC-contlg
US-09-078-294-4

Query Match 49.6%; Score 248.4; DB 4; Length 80246;
Best Local Similarity 72.8%; Pred. No. 9.2e-64;
Matches 361; Conservative 0; Mismatches 131; Indels 4; Gaps 3;

OY	3	TTTGSGTGTGATGACATGAGAGTCCGTCGCCATCCATGCGCCGAATGGTATGGCTGGG	62
Db	49139	TTTGGATTTTAAATAAATAATCTCTTCCTCAAGGCTATGCCCAAGAACACATTTTCTAGG	49198
OY	63	TTTTCTCTAAGGATTTATAGTTTATAGTCTAACATTTAAAGTCTTTAAATCCATCTGAA	122
Db	49199	TTTTCTCTAAGGATTTCTATAGTGTCAAAAGCTTAATTTAAAGCTTTTAAATCCACCTCAAG	49258
OY	123	TTAATTTTATATAAGGTGTAAAGGAGATCCAGTTTACAGTCTTCTAATAGGCGCTAC	182
Db	49259	TTAATTTTATATAAGGTGTAAAGGAGATCCAGTCTTCTAATAGGCGCTAC	49318
OY	183	CAGTTTTCTCAGCACCATTTATTAATAATAGGAATCCCTTCCCATCTGTTGTTTCTCA	242
Db	49319	CAGCAATCCAGAACCATTTATTTGAATAAAGAAATCTTTCCCATTTGCTT - ATTTTGTCA	49377
OY	243	GGTTTGTCAAAAGATCAGATAGTGTGTAGATATGCGGCAATATTTCTAGAGGCTCTGTTCTG	302
Db	49378	ACTTTGTCAAAAGATCGGATAGCTAGTGAAGAGTGGG - TTTTTCCTGGTGTATCTACTCG	49435
OY	303	TTCCATTTGTTGATATCTCTGTTTGGTACCAATACATGTTGTTGGTACTGAGCC	362
Db	49436	TTACATTTGGCTATAGTGTCTGTTTGTATACATATACATGCGTTTTGTACTATAGTGC	49495
OY	363	TTTGAAGCTATGTTGAAGTACAGTACATGATGCCCTCCAGCTTTGTTCTTTTGGCTTAG	422
Db	49496	TCATTAACATATGTTAAAGTGGATAATATGTAATCCCT - CTGCTTGGCTGTTTGGCTTAG	49554
OY	423	ATTGACTTGGCGATAGTGGGCTCTTTTGTGTTCTCAATCAATCAATTTAACTAGTTTTTTC	482
Db	49555	ATTGCTTTGGCTATGAGGCTCTCTTTTTCACCTTCAATATGAAATTTTGAATAGTTTTTCT	49614
OY	483	AATTCGTGTGAAGAAG 498	
Db	49615	AATTCCTTGAATAATG 49630	

[illegible]

OY	183	CAGTTTTCGAGCCACCTTATTAATAANAGGAAATCCCTTCCCATGCTGTTTGTTCGA	242
Db	49581	CAGCAATCCCAAGAACCTTTATTGAATAAAGCAATCTTTCCTATTGGCT - ATTGTTGA	49639
OY	243	GGTTTGCAAAAGATCATAGTATTGTAGATATGAGGCAATATTTCTAGAGGCTCTGTCTGA	302
Db	49640	ACTTTGTCMAAGATCGGATGACGTAGAGACTGTGGC - TTTTCTGTGGTTATCTACTCTG	49697
OY	303	TTCCATTGGTTGATATCTCTGTTTGGTATACAGTACCATGTTGTTTGGTACTGTAGCC	362
Db	49698	TTCAATTGGCTATGCTGTCTGTTTGTATCAATATCAATGCTGTTTGTACTATAGTGC	49757
OY	363	TTGTAGTGTAGTTTGAAGTCAGGTAGCATGATGCTCCACGTTTGTCTTTGGCTTAG	422
Db	49758	TCATACATAGTTTAAAGTGTGAATTAAGTATGCT - CTGCTTCTGCTGTTTGGCTTAG	49816
OY	423	ATTACACTGGGATGTGGGCTCTTTTGGTTTCCACATGAACCTTAAAGTACTGTTTTTGC	482
Db	49817	ATTGCTTTGGCTATGAGGCTCTTTTTCCTACTTCATATGAATTTTGAATAGTTTTTTCT	49876
OY	483	AATCTGGAAGAAG 498	
Db	49877	AATCTTTGAAAAATG 49892	

QY	1	CTTTTGGTGTGTTTACATGAGTCCTGTCGCCATCGCTATGTCGTAATGTAATGCTCG	60
DB	31100	CTTTTGAGGACTGTAATGTAATTAATTATTACCTGGCCAAATGGCCAGAGATATTCTTA	310411
QY	61	GGTTTCTCTAGGAGATTTATGTTTTAGGCTCAACATTTAAGTCTTTAATCACTCTGG	120
DB	31040	AGTTTGATCTAGGATTTTTTATATGTTATGTCCTTCACTTAAAGTCCTTATTTCCACCTGG	309811
QY	121	AATTAATTTTGTATAGGTGTGAAGGAGAGGATCCAGTTTTCACCTTTCTACATAGGGCTA	180
DB	30980	TGTTAATTTTGTATAGGTGTGAAGGAGAGGATTTGATTTCACTCTTCGCAATAGGTTA	309211
QY	181	GCCAGTTTCTCAGCACACCATTTTATTAATAGGCAATCT-----TTCCCATTTGCTT	232
DB	30920	GCCAGTTTCTCAGCACACCATTTTATGATGGGTGTCCTAATGGTATTTCCCATTTGCTT	308611
QY	233	GTTTTCCTCAGGTTTGCAAGAATGATGTTGTGATATGGGAGATATTCTGAGGG	292
DB	30860	ATTTTGTCCACTTGTGTAAGATCACTTGTGTAGGTGTACAGTTTATTTTCAGGGGT	308011
QY	293	CTCTGTTCTGTCCATTTGTTGATATCTCG-----TTTTGGTACAGATACATGTTGT	346

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 15:04:06 ; Search time 166.778 Seconds

(without alignments)
6765.003 Million cell updates/sec

Title: US-10-083-853b-2_COPY_29421_29921

Perfect score: 501
Sequence: 1 ctttgcgttttagacatgcattctgtgaagaagaaatga 501

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database 8
1: /SID2/gcgdata/geneseq/geneqn-emb1/NA1980.DAT:*
2: /SID2/gcgdata/geneseq/geneqn-emb1/NA1981.DAT:*
3: /SID2/gcgdata/geneseq/geneqn-emb1/NA1982.DAT:*
4: /SID2/gcgdata/geneseq/geneqn-emb1/NA1983.DAT:*
5: /SID2/gcgdata/geneseq/geneqn-emb1/NA1984.DAT:*
6: /SID2/gcgdata/geneseq/geneqn-emb1/NA1985.DAT:*
7: /SID2/gcgdata/geneseq/geneqn-emb1/NA1986.DAT:*
8: /SID2/gcgdata/geneseq/geneqn-emb1/NA1987.DAT:*
9: /SID2/gcgdata/geneseq/geneqn-emb1/NA1988.DAT:*
10: /SID2/gcgdata/geneseq/geneqn-emb1/NA1989.DAT:*
11: /SID2/gcgdata/geneseq/geneqn-emb1/NA1990.DAT:*
12: /SID2/gcgdata/geneseq/geneqn-emb1/NA1991.DAT:*
13: /SID2/gcgdata/geneseq/geneqn-emb1/NA1992.DAT:*
14: /SID2/gcgdata/geneseq/geneqn-emb1/NA1993.DAT:*
15: /SID2/gcgdata/geneseq/geneqn-emb1/NA1994.DAT:*
16: /SID2/gcgdata/geneseq/geneqn-emb1/NA1995.DAT:*
17: /SID2/gcgdata/geneseq/geneqn-emb1/NA1996.DAT:*
18: /SID2/gcgdata/geneseq/geneqn-emb1/NA1997.DAT:*
19: /SID2/gcgdata/geneseq/geneqn-emb1/NA1998.DAT:*
20: /SID2/gcgdata/geneseq/geneqn-emb1/NA1999.DAT:*
21: /SID2/gcgdata/geneseq/geneqn-emb1/NA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneqn-emb1/NA2001A.DAT:*
23: /SID2/gcgdata/geneseq/geneqn-emb1/NA2001B.DAT:*
24: /SID2/gcgdata/geneseq/geneqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	478.6	95.5	5065	21 AAC69111	Human secreted pro
2	478.6	95.5	5065	22 AAC69111	Human secreted pro
3	477.4	95.3	4709	21 AAC69112	Human secreted pro
4	473.8	94.6	1542	23 AAS67689	DNA encoding novel
5	473.8	94.6	1542	23 AAS67689	DNA encoding novel
6	473.8	94.6	1758	23 AAS74901	DNA encoding novel
7	473.8	94.6	1758	23 AAS75196	DNA encoding novel
8	473.8	94.6	1942	23 AAS80525	DNA encoding novel
9	473.8	94.6	2205	23 AAS72639	DNA encoding novel

C 10	473.8	94.6	3052	23 AAS76197	DNA encoding novel
C 11	473.8	94.6	3105	23 AAS78724	DNA encoding novel
C 12	473.8	94.6	3105	23 AAS78792	DNA encoding novel
C 13	473.8	94.6	3105	23 AAS79645	DNA encoding novel
C 14	473.8	94.6	3239	23 AAS76858	DNA encoding novel
C 15	473.8	94.6	4114	23 AAS77062	DNA encoding novel
C 16	473.8	94.6	4289	23 AAS75564	DNA encoding novel
C 17	473.8	94.6	4502	23 AAS78636	DNA encoding novel
C 18	473.8	94.6	5142	23 AAS68835	DNA encoding novel
C 19	473.8	94.6	5654	23 AAS76524	DNA encoding novel
C 20	473.8	94.6	8779	23 AAS71656	DNA encoding novel
C 21	473.8	94.6	13931	23 AAS72644	DNA encoding novel
C 22	472.2	94.3	1827	23 AAS76723	DNA encoding novel
C 23	472.2	94.3	1096	23 AAS76063	DNA encoding novel
C 24	472.2	94.3	1197	22 AAK77110	Human immune/haema
C 25	472.2	94.3	1200	22 AAK77109	Human immune/haema
C 26	472.2	94.3	1248	23 AAS69649	DNA encoding novel
C 27	472.2	94.3	1248	23 AAS74285	DNA encoding novel
C 28	472.2	94.3	1248	23 AAS76231	DNA encoding novel
C 29	472.2	94.3	1569	23 AAS68463	DNA encoding novel
C 30	472.2	94.3	1569	23 AAS71273	DNA encoding novel
C 31	472.2	94.3	1821	23 AAS80517	DNA encoding novel
C 32	472.2	94.3	2365	23 AAS73523	DNA encoding novel
C 33	472.2	94.3	2606	23 AAS78958	DNA encoding novel
C 34	472.2	94.3	2940	23 AAS78729	DNA encoding novel
C 35	472.2	94.3	3341	23 AAS71209	DNA encoding novel
C 36	472.2	94.3	3471	23 AAS68587	DNA encoding novel
C 37	472.2	94.3	3579	23 AAS70252	DNA encoding novel
C 38	472.2	94.3	3579	23 AAS73833	DNA encoding novel
C 39	472.2	94.3	4555	23 AAS84104	DNA encoding novel
C 40	472.2	94.3	6334	23 AAS76938	DNA encoding novel
C 41	472.2	94.3	30274	23 AAS85251	DNA encoding novel
C 42	472.2	94.3	167343	24 AB164403	Stomach cancer rel
C 43	472.2	94.3	167343	24 AB167239	Thyroid cancer rel
C 44	470.6	93.9	1200	23 AAS76061	DNA encoding novel
C 45	470.6	93.9	1200	23 AAS76691	DNA encoding novel

ALIGNMENTS

RESULT 1	
ID AAC69111	standard; DNA; 5065 BP.
AC AAC69111;	
XX	
XX	31-JAN-2001 (first entry)
XX	
DE	Human secreted protein gene 3 clone HPJCL13.
XX	
KW	Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW	antiallergic; hepatotropic; antidiabetic; antiinflammatory; anticancer;
KW	vulnerable; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW	cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW	neurological disease; infection; human; secreted protein; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200055371-A1.
XX	
PD	21-SEP-2000.
XX	
PF	16-MAR-2000; 2000WO-US06783.
XX	
PR	18-MAR-1999; 99US-0125055.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	
PI	Ruben SM, Ni J, Ebner R, Rosen CA, Shi Y, Birse C, Florence K;
PI	Komatsu S, G, Lafleur DW, Moore PA, Olsen HS, Young PE;
XX	
DR	WPI; 2000-594448/56.

Best Local Similarity 95.4%; Pred. No. 1.5e-120;
Matches 478; Conservative 12; Mismatches 11; Indels 0; Gaps 0;

```
OY 1 CTTTGGCTTTAGACATGAGTCCTTCCCATGCCATATGCTGATGCTGCTG 60
    |||||||
DB 1090 CTTTGGCTTTAGACATGAGTCCTTCCCATGCCATATGCTGATGCTGCTG 1149
    |||||||
OY 61 GGTTCCTTACAGGATTTATGCTTTAGTCTTAACTTAAAGCTTTAACTTCTG 120
    |||||||
DB 1150 GGTTCCTTACAGGATTTATGCTTTAGTCTTAACTTAAAGCTTTAACTTCTG 1209
    |||||||
OY 121 AATTAAATTTTGTATTAAGCTGTAAAGAGGATCCAGTTTCACTTTACATAGGCTA 180
    |||||||
DB 1210 AATTAAATTTTGTATTAAGCTGTAAAGAGGATCCAGTTTCACTTTACATAGGCTA 1269
    |||||||
OY 181 GCCAGTTTTCACACACCATTTATTAATTAAGGAATCCCTTCCCATGCTTGTCTTCT 240
    |||||||
DB 1270 GCCAGTTTTCACACACCATTTATTAATTAAGGAATCCCTTCCCATGCTTGTCTTCT 1329
    |||||||
OY 241 CAGGTTTGTCAAAAGATCAGATAGTTGTAGATGCGGCAATTTTCTGAGGGCTCTGTC 300
    |||||||
DB 1330 CAGGTTTGTCAAAAGATCAGATAGTTGTAGATGCGGCAATTTTCTGAGGGCTCTGTC 1389
    |||||||
OY 301 TGTTCATTTGTTGATATCTCTGTTTGGTACCAAGTACCAATGTTTGGTTACTGTAG 360
    |||||||
DB 1390 TGTTCATTTGTTGATATCTCTGTTTGGTACCAAGTACCAATGTTTGGTTACTGTAG 1449
    |||||||
OY 361 CCTGTGATGCTAGTTTGAAGTACAGTACGATGATGCCCTCCAGCTTGTCTTTGGCTTA 420
    |||||||
DB 1450 CCTGTGATGCTAGTTTGAAGTACAGTACGATGATGCCCTCCAGCTTGTCTTTGGCTTA 1509
    |||||||
OY 421 GGATTTGACTTGCGCATGCGGCTCTTTTGGTCCACATCACTTTAAAGTATGTTT 480
    |||||||
DB 1510 GGATTTGACTTGCGCATGCGGCTCTTTTGGTCCACATCACTTTAAAGTATGTTT 1569
    |||||||
OY 481 CCAATTTCTGTGAAGAACTAA 501
    |||||||
DB 1570 CCAATTTCTGTGAAGAACTAA 1590
    |||||||

RESULT 3
AAC69112/c
ID AAC69112 standard; DNA; 4709 BP.
AC AAC69112;
XX 31-JAN-2001 (first entry)
DE Human secreted protein gene 3 clone HNHCT15.
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antileuker;
KW vulnerrary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardial; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein; ss.
XX Homo sapiens.
OS
XX WO200055371-A1.
PN 21-SEP-2000.
PD 16-MAR-2000; 2000WO-US06783.
PF 18-MAR-1999; 99US-0125055.
PR
XX (HUMA-) HUMAN GENOME SCI INC.
PA Ruben SM, Ni J, Edner R, Rosen CA, Shi Y, Birse C, Florence K;
PI Komatsoulis G, Lafleur DW, Moore PA, Olsen HS, Young PE;
XX WPI: 2000-594448/56.
DR P-PSDB; AAB38012.
```

XX New nucleic acid molecules encoding 27 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PI used as food additives or preservatives -

PS Claim 1; Page 385-386; 453pp; English.

XX The invention relate to the isolation of genes AAC69084-C69119 encoding
CC 27 human secreted proteins AAB37984-B38019. The genes can be used to
CC generate fusion proteins by linking to the gene for the human
CC immunoglobulin G Fc portion (AAC69075) for increasing the stability of
CC the fusion protein as compared to the human protein only. The genes and
CC proteins are useful for preventing, ameliorating or treating medical
CC conditions, e.g. by protein or gene therapy. The genes are isolated
CC from a range of human tissues disclosed in the specification. The
CC nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer, and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d)
CC wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.

XX Sequence 4709 BP; 1868 A; 990 C; 836 G; 939 T; 76 other;

XX Query Match 95.3%; Score 477.4; DB 21; Length 4709;

XX Best Local Similarity 95.4%; Pred. No. 3.2e-120;

XX Matches 478; Conservative 11; Mismatches 12; Indels 0; Gaps 0;

```
OY 1 CTTTGGCTTTAGACATGAGTCCTTCCCATGCCATATGCTGATGCTGCTG 60
    |||||||
DB 3654 CTTTGGCTTTAGACATGAGTCCTTCCCATGCCATATGCTGATGCTGCTG 3595
    |||||||
OY 61 GGTTCCTTACAGGATTTATGCTTTAGTCTTAACTTAAAGCTTTAACTTCTG 120
    |||||||
DB 3594 GGTTCCTTACAGGATTTATGCTTTAGTCTTAACTTAAAGCTTTAACTTCTG 3535
    |||||||
OY 121 AATTAAATTTTGTATTAAGCTGTAAAGAGGATCCAGTTTCACTTTACATAGGCTA 180
    |||||||
DB 3534 AATTAAATTTTGTATTAAGCTGTAAAGAGGATCCAGTTTCACTTTACATAGGCTA 3475
    |||||||
OY 181 GCCAGTTTTCACACACCATTTATTAATTAAGGAATCCCTTCCCATGCTTGTCTTCT 240
    |||||||
DB 3474 GCCAGTTTTCACACACCATTTATTAATTAAGGAATCCCTTCCCATGCTTGTCTTCT 3415
    |||||||
OY 241 CAGGTTTGTCAAAAGATCAGATAGTTGTAGATGCGGCAATTTTCTGAGGGCTCTGTC 300
    |||||||
DB 3414 CAGGTTTGTCAAAAGATCAGATAGTTGTAGATGCGGCAATTTTCTGAGGGCTCTGTC 3355
    |||||||
OY 301 TGTTCATTTGTTGATATCTCTGTTTGGTACCAAGTACCAATGTTTGGTTACTGTAG 360
    |||||||
DB 3354 TGTTCATTTGTTGATATCTCTGTTTGGTACCAAGTACCAATGTTTGGTTACTGTAG 3295
    |||||||
OY 361 CCTTCTAGTGTAGTTTGAAGTACAGTACGATGATGCCCTCCAGCTTGTCTTTGGCTTA 420
    |||||||
DB 3294 CCTTCTAGTGTAGTTTGAAGTACAGTACGATGATGCCCTCCAGCTTGTCTTTGGCTTA 3235
    |||||||
OY 421 GGATTTGACTTGCGCATGCGGCTCTTTTGGTCCACATCACTTTAAAGTATGTTT 480
    |||||||
DB 3234 GGATTTGACTTGCGCATGCGGCTCTTTTGGTCCACATCACTTTAAAGTATGTTT 3175
    |||||||
OY 481 CCAATTTCTGTGAAGAACTAA 501
    |||||||
DB 3174 CCAATTTCTGTGAAGAACTAA 3154
    |||||||

RESULT 4
AAS67689/c
ID AAS67689 standard; cDNA; 1542 BP.
```

XX AAS67689;
AC 13-FEB-2002 (first entry)
DT
XX
DE DNA encoding novel human diagnostic protein #3493.
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
PN WO200175067-A2.
PD 11-OCT-2001.
PE 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
PI Dmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
DR P-PSDB; ABG03502.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
XX Claim 1; SEQ ID No 3493; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX

Sequence 1542 BP; 629 A; 304 C; 303 G; 306 T; 0 other;

Query Match 94.6%; Score 473.8; DB 23; Length 1542;
Best Local Similarity 96.6%; Pred. No. 2.1e-119;
Matches 484; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 1 CTTTGGTGTAGACATGAGTCCTGCGCATGCCATGCTCCGATGATGCGCTG 60
DB 1432 CTTTGGGTTGAGACATGAGTCCTGCGCATGCCATGCTCCGATGATGCGCTA 1373
QY 61 GGTTCCTCTAGGAGATTATGTTAGGCTTAACATTTAAGCTTTTAATCCATCTTG 120
DB 1372 GGTTCCTCTAGGAGATTATGTTAGGCTTAACATTTAAGCTTTTAATCCATCTTG 1313
QY 121 AATTAATTTTGTATAGGTATAGGAAGGATCCAGTTTACGTTCTTAATAGGCTA 180
|||||

DB 1312 AATTAATTTTGTATAGGTATAGGAAGGATCCAGTTTACGTTCTTAATAGGCTA 1253
QY 181 GCCAGTTTCTCAGCACCATTATTAATAGGAAATCCTTCCCATTCGTTGTTTCT 240
DB 1252 GCCAGTTTCTCAGCACCATTATTAATAGGAAATCCTTCCCATTCGTTGTTTCT 1193
QY 241 CAGGTTTGTCAAAAGATGATAGTTGATATAGGCGGCAATTTCTGAGGCGCTGTC 300
DB 1192 CAGGTTTGTCAAAAGATGATAGTTGATATAGGCGGCAATTTCTGAGGCGCTGTC 1133
QY 301 TGTTCACATGTTGATATATCTGTTTGGTCCACGATACGATGTTTGGTTACTGTAG 360
DB 1132 TGTTCACATGTTGATATATCTGTTTGGTCCACGATACGATGTTTGGTTACTGTAG 1073
QY 361 CCTGTAGTATAGTTTGAATGATCAGATAGATGCTCCAGCTTTGTTTGGCTTA 420
DB 1072 CCTGTAGTATAGTTTGAATGATCAGATAGATGCTCCAGCTTTGTTTGGCTTA 1013
QY 421 GGATTGACTTGGCAGATGTGGGCTCTTTTGGTCCACATGAACTTTAAAGTATTTT 480
DB 1012 GGATTGACTTGGCAGATGTGGGCTCTTTTGGTCCACATGAACTTTAAAGTATTTT 953
QY 481 CCAATTCGTGAAGAAAGTAA 501
DB 952 CCAATTCGTGAAGAAAGTCA 932

RESULT 5

AAS76725/C
ID AAS76725 standard; CDNA; 1542 BP.

AAS76725;

13-FEB-2002 (first entry)

DNA encoding novel human diagnostic protein #12529.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PE 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

PI Dmanac RT, Liu C, Tang YT;

XX WPI: 2001-639362/73.

DR P-PSDB; ABG12538.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity

XX Claim 1; SEQ ID No 12529; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX

SO Sequence 1542 BP; 629 A; 304 C; 303 G; 306 T; 0 other;

Query Match 94.6%; Score 473.8; DB 23; Length 1542;
Best Local Similarity 96.6%; Pred. No. 2.1e-119;
Matches 484; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

```
OY 1 CTTTGGTGTAGACATGAGTCCCTGGCCATGCTATGCTGATGATGCTG 60
DB 1432 CTTTGGTGTAGACATGAGTCCCTGGCCATGCTATGCTGATGATGCTG 1373
OY 61 GGTTCCTCTAGAGATTTTATGTTTATAGCTTAACATTTAATTCATCTTG 120
DB 1372 GGTTCCTCTAGAGATTTTATGTTTATAGCTTAACATTTAATTCATCTTG 1313
OY 121 AATTAAATTTTGTATAGGTGTAAGGAGGATCCAGTTTCACTTTCAATAGGCTTA 180
DB 1312 AATTAAATTTTGTATAGGTGTAAGGAGGATCCAGTTTCACTTTCAATAGGCTTA 1253
OY 181 GCCAGTTTCTCAGCACCATTATTAATAGGAATCCTTCCCATATGCTTTTCTT 240
DB 1252 GCCAGTTTCTCAGCACCATTATTAATAGGAATCCTTCCCATATGCTTTTCTT 1193
OY 241 CAGTTTGTCAAGATCAGATTAATGATGAGGATATTTTCTGAGGCTCTGTTT 300
DB 1192 CAGTTTGTCAAGATCAGATTAATGATGAGGATATTTTCTGAGGCTCTGTTT 1133
OY 301 TGTTCATTTGTTGATATCTCTGTTTGGTACAGTACATGTTTGGTACGTAG 360
DB 1132 TGTTCATTTGTTGATATCTCTGTTTGGTACAGTACATGTTTGGTACGTAG 1073
OY 361 CCTGTAGTGTAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 420
DB 1072 CCTGTAGTGTAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1013
OY 421 GGATGACTTGGGATGTGGCTTTTGGTTCACATGAATTAAGTATGTTT 480
DB 1012 GGATGACTTGGGATGTGGCTTTTGGTTCACATGAATTAAGTATGTTT 953
OY 481 CCAATTCCTGGAAGAAAGTAA 501
DB 952 CCAATTCCTGGAAGAAAGTAA 932
```

RESULT 6

AAS74901/C
ID AAS74901 standard; cDNA; 1758 BP.

AC AAS74901;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #10705.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

```
XX 11-OCT-2001.
PD 30-MAR-2001; 2001MO-US08631.
XX 31-MAR-2000; 2000US-0540217.
PF 23-AUG-2000; 2000US-0649167.
PR (HYSE-) HYSEQ INC.
XX Dmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
DR P-PSDB; ABG10714.
XX
```

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity

PS Claim 1; SEQ ID NO 10705; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX

SO Sequence 1758 BP; 676 A; 381 C; 332 G; 369 T; 0 other;

Query Match 94.6%; Score 473.8; DB 23; Length 1758;
Best Local Similarity 96.6%; Pred. No. 2.2e-119;
Matches 484; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

```
OY 1 CTTTGGTGTAGACATGAGTCCCTGGCCATGCTATGCTGATGATGCTG 60
DB 1159 CTTTGGTGTAGACATGAGTCCCTGGCCATGCTATGCTGATGATGCTG 1100
OY 61 GGTTCCTCTAGAGATTTTATGTTTATAGCTTAACATTTAATTCATCTTG 120
DB 1099 GGTTCCTCTAGAGATTTTATGTTTATAGCTTAACATTTAATTCATCTTG 1040
OY 121 AATTAAATTTTGTATAGGTGTAAGGAGGATCCAGTTTCACTTTCAATAGGCTTA 180
DB 1039 AATTAAATTTTGTATAGGTGTAAGGAGGATCCAGTTTCACTTTCAATAGGCTTA 980
OY 181 GCCAGTTTCTCAGCACCATTATTAATAGGAATCCTTCCCATATGCTTTTCTT 240
DB 979 GCCAGTTTCTCAGCACCATTATTAATAGGAATCCTTCCCATATGCTTTTCTT 920
OY 241 CAGTTTGTCAAGATCAGATTAATGATGAGGATATTTTCTGAGGCTCTGTTT 300
DB 919 CAGTTTGTCAAGATCAGATTAATGATGAGGATATTTTCTGAGGCTCTGTTT 860
OY 301 TGTTCATTTGTTGATATCTCTGTTTGGTACAGTACATGTTTGGTACGTAG 360
DB 859 TGTTCATTTGTTGATATCTCTGTTTGGTACAGTACATGTTTGGTACGTAG 800
```


DR P-PSDB: ABG16338.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 1; SEQ ID No 16329; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II), or to treat disease states involving
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 1942 BP; 754 A; 419 C; 368 G; 401 T; 0 other;

Query Match 94.6%; Score 473.8; DB 23; Length 1942;
Best Local Similarity 96.6%; Pred. No. 2.3e-119;
Matches 484; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

OY 1 CTTTGTGCTTTAGACATGAAGTCCCTGGCCATGCTATGCTGAATGGTATGCTG 60
DB 1343 CTTTGTGCTTTAGACATGAAGTCCCTGGCCATGCTATGCTGAATGGTATGCTG 1284
OY 61 GGTTCCTCTAGAGATTTTANGGTTTAAAGTCTAACATTTAAGCTTTAATCATCTTG 120
DB 1283 GGTTCCTCTAGAGATTTTANGGTTTAAAGTCTAACATTTAAGCTTTAATCATCTTG 1224
OY 121 AATTATTTTGTATAGGTATAGAGAGGATCCAGTTCACCTTCTCATAGGCTTA 180
DB 1223 AATTATTTTGTATAGGTATAGAGAGGATCCAGTTCACCTTCTCATAGGCTTA 1164
OY 181 GCCAGTTTCTCAGACCATTTTAAATAGGGAATCCTTCCCATTTGCTTTTCT 240
DB 1163 GCCAGTTTCTCAGACCATTTTAAATAGGGAATCCTTCCCATTTGCTTTTCT 1104
OY 241 CAGGTTTGTCAAGATCATAGTGTATAGATGCGCATTTTCTGAGGCTCTGTTTC 300
DB 1103 CAGGTTTGTCAAGATCATAGTGTATAGATGCGCATTTTCTGAGGCTCTGTTTC 1044
OY 301 TGTTCATGCTGATATCTGTTTGTACACAGATCATGTTGTTTGGTACTGTAG 360
DB 1043 TGTTCATGCTGATATCTGTTTGTACACAGATCATGTTGTTTGGTACTGTAG 994
OY 361 CTTTGTAGTGTAGTGTAGATGAGTATGAGTCCATGCTTGTCTTTGGCTTA 420
DB 983 CTTTGTAGTGTAGTGTAGATGAGTATGAGTCCATGCTTGTCTTTGGCTTA 924
OY 421 GGATTGACTTGGCGAGTGGGCTCTTTTGGTTCACATGAACCTTAAGTATGTTT 480
DB 923 GGATTGACTTGGCGAGTGGGCTCTTTTGGTTCACATGAACCTTAAGTATGTTT 864
OY 481 CCAATTCGTGGAAGAAGTAA 501
DB 863 CCAATTCGTGGAAGAAGTAA 843

RESULT 9
AAS72639/c
ID AAS72639 standard; cDNA; 2205 BP.
XX
XX AAS72639;
AC
XX
XX 13-FEB-2002 (first entry)
DT
XX
DE DNA encoding novel human diagnostic protein #8443.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
OS
XX
XX W0200175067-A2.
PN
XX
PD 11-OCT-2001.
XX
XX 30-MAR-2001; 2001MO-US08631.
PF
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Dmanac RT, Liu C, Tang YT;
PI
XX
XX WPI: 2001-639362/73.
DR
XX
DR P-PSDB: ABC08452.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1; SEQ ID No 8443; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II), or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 2205 BP; 928 A; 441 C; 396 G; 440 T; 0 other;

Query Match 94.6%; Score 473.8; DB 23; Length 2205;
Best Local Similarity 96.6%; Pred. No. 2.4e-119;
Matches 484; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

OY 1 CTTTGTGCTTTAGACATGAAGTCCCTGGCCATGCTATGCTGAATGGTATGCTG 60
DB 1378 CTTTGTGCTTTAGACATGAAGTCCCTGGCCATGCTATGCTGAATGGTATGCTG 1319
OY 61 GGTTCCTCTAGAGATTTTANGGTTTAAAGTCTAACATTTAAGCTTTAATCATCTTG 120
DB 1378 CTTTGTGCTTTAGACATGAAGTCCCTGGCCATGCTATGCTGAATGGTATGCTG 1319

DB 1318 GGTTCCTCTTACGGTTTATAGCTTTAGCTACGTTTAACTTTATTCATCTTG 1259
QY 121 AATTATTTTGTATAGGTATAGGAAGGATCCAGTTTACGTTTCTACATAGGCTTA 180
DB 1258 AATTATTTTGTATAGGTATAGGAAGGATCCATTTTACGTTTCTACATAGGCTTA 1199
QY 181 GCCAGTTTCTCAGACCATTTATTAATAGGAATCCCTTCCCATTCCTGTTTCT 240
DB 1198 GCCAGTTTCTCAGACCATTTATTAATAGGAATCCCTTCCCATTCCTGTTTCT 1139
QY 241 CAGTTTGTCAAAAGATCAGATAGTGTAGATATGGGCAATTTTCTGAGGGCTGTTTC 300
DB 1138 CAGTTTGTCAAAAGATCAGATAGTGTAGATATGGGCAATTTTCTGAGGGCTGTTTC 1079
QY 301 TGTTCATTTGTTGATATCTCTGTTTGGTACACATGATGTTTGGTGTACGTAG 360
DB 1078 TGTTCATTTGATATCTCTGTTTGGTACACATGATGTTTGGTGTACGTAG 1019
QY 361 CCTGTAGTGTATAGTGTAGATATGGGCAATTTTCTGAGGGCTGTTTC 420
DB 1018 CCTGTAGTGTATAGTGTAGATATGGGCAATTTTCTGAGGGCTGTTTC 959
QY 421 GGATTTGACTGGCGATGCGCTCTTTTGGTCCACATGAATTTAAAGTATTTT 480
DB 958 GGATTTGACTGGCGCGAGCGGCTCTTTTGGTCCACATGAATTTAAAGTATTTT 899
QY 481 CCAATTCGTGTGAAGAACTAA 501
DB 898 CCAATTCGTGTGAAGAACTAA 878

RESULT 10

AA576197/c
ID AA576197 standard; cDNA; 3052 BP.

AC AA576197;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #12001.

KM Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR P-PSDB; ABG12010.

XX New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity -

PS Claim 1: SEQ ID No 12001; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AA576197-AA594564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 3052 BP; 1044 A; 704 C; 639 G; 665 T; 0 other;

Query Match 94.6%; Score 473.8; DB 23; Length 3052;

Best Local Similarity 96.6%; Pred. No. 2.7e-119;

Matches 484; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 1 CTTTGGTCTTTTAAACATGAAGTCTCCCATCCCATGCTGATGATGCTG 60

DB 1717 CTTTGGTCTTTTAAACATGAAGTCTCCCATCCCATGCTGATGATGCTG 1658

QY 61 GGTTTCTCTTCTAGGATTTATGTTTATGCTTACATTTTAACTTCACTCTG 120

DB 1657 GGTTTCTCTTCTAGGATTTATGTTTATGCTTACATTTTAACTTCACTCTG 1598

QY 121 AATTATTTTGTATAGGTATAGGAAGGATCCAGTTTACGTTTCTACATAGGCTTA 180

DB 1597 AATTATTTTGTATAGGTATAGGAAGGATCCAGTTTACGTTTCTACATAGGCTTA 1538

QY 181 GCCAGTTTCTCAGACCATTTATTAATAGGAATCCCTTCCCATGCTGTTTCT 240

DB 1537 GCCAGTTTCTCAGACCATTTATTAATAGGAATCCCTTCCCATGCTGTTTCT 1478

QY 241 CAGTTTGTCAAAAGATCAGATAGTGTAGATATGGGCAATTTCTGAGGGCTCTGTC 300

DB 1477 CAGTTTGTCAAAAGATCAGATAGTGTAGATATGGGCAATTTCTGAGGGCTCTGTC 1418

QY 301 TGTTCATTTGTTGATATCTCTGTTTGGTACACATGCTGTTTGGTACTGTAG 360

DB 1417 TGTTCATTTGTTGATATCTCTGTTTGGTACACATGCTGTTTGGTACTGTAG 1358

QY 361 CCTTGTAGTGTATAGGTATAGGAAGGATCCCTTCCCATGCTGTTTGGCTTA 420

DB 1357 CCTTGTAGTGTATAGGTATAGGAAGGATCCCTTCCCATGCTGTTTGGCTTA 1298

QY 421 GGATTTGACTGGCGATGCGCTCTTTTGGTCCACATGAATTTTAAAGTATTTT 480

DB 1297 AGATTGACTTGGCAATGCGGCTCTTTTGGTCCACATGAATTTTAAAGTATTTT 1238

QY 481 CCAATTCGTGTGAAGAACTAA 501

DB 1237 CCAATTCGTGTGAAGAACTAA 1217

RESULT 11

AA578724/c
ID AA578724 standard; cDNA; 3105 BP.

AC AA578724;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #14528.

KM Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX OS Homo sapiens.
 XX PN WO200175067-A2.
 XX 11-OCT-2001.
 XX PF 30-MAR-2001; 2001WO-US08631.
 XX PR 31-MAR-2000; 2000US-0540217.
 XX PR 23-AUG-2000; 2000US-0649167.
 XX PA (HYSE-) HYSEQ INC.
 XX PI Drmanac RT, Liu C, Tang YT;
 XX DR WPI; 2001-639362/73.
 XX DR P-PSDB; ABG14537.
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX Claim 1; SEQ ID NO 14528; 103bp; English.
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC and responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcr_sequences.
 XX Sequence 3105 BP; 1067 A; 721 C; 658 G; 659 T; 0 other;
 SO
 Query Match 94.6%; Score 473.8; DB 23; Length 3105;
 Best Local Similarity 96.6%; Pred. No. 2.7e-119;
 Matches 484; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 1 CTTTGGTGTGTTAGACATGAAGTCCTTGCCCATGCTCATGCTGTAATGATTCCTG 60
 DB 1717 CTTTGGTGTGTTAGACATGAAGTCCTTGCCCATGCTCATGCTGTAATGATTCCTG 1658
 QY 61 GGTTCCTCTAGGAGATTTATGTTTAAAGTCTTAACCTTTAATCCATCTTG 120
 DB 1657 GGTTCCTCTAGGAGATTTATGTTTAAAGTCTTAACCTTTAATCCATCTTG 1598
 QY 121 AATTAAATTTGTATAGGTGAAGGAAGGATCCAGTTTCAAGTTTCTACATAGGCTA 180
 DB 1597 AATTAAATTTGTATAGGTGAAGGAAGGATCCAGTTTCAAGTTTCTACATAGGCTA 1538
 QY 181 GCCAGTTTCTCAGCACCATTTAATTAATAGGAATCCCTTCCCATGCTGTTTCT 240
 DB 1537 GCCAGTTTCTCAGCACCATTTAATTAATAGGAATCCCTTCCCATGCTGTTTGT 1478
 QY 241 CAGGTTGTCAAAAGATCAGATAGTTGATATGCGGATATTTCTAGAGGCTGTTC 300
 DB 1477 CAGGTTGTCAAAAGATCAGATAGTTGATATGCGGATATTTCTAGAGGCTGTTC 1418

QY 301 TGTTCATGTTGATATCTCTGTTTGGTACAGTACATGTTGTTGTTACTGTAG 360
 DB 1417 TGTTCATGTTGATATCTCTGTTTGGTACAGTACATGTTGTTGTTACTGTAG 1358
 QY 361 CCTTGATGTTAGTTTGAAGTCAGTACATGATCCCTCCAGCTTGTCTTTGGCTTA 420
 DB 1357 CCTTGATGTTAGTTTGAAGTCAGTACATGATCCCTCCAGCTTGTCTTTGGCTTA 1298
 QY 421 CGATTGACTTGGCGATGCGCTCTTTTGGTTCCATGATGACTTTAAAGTATTTT 480
 DB 1297 AGATTGACTTGGCGATGCGCTCTTTTGGTTCCATGATGACTTTAAAGTATTTT 1238
 QY 481 CCAATTCTGTGAAGAAAGTAA 501
 DB 1237 CCAATTCTGTGAAGAAAGTCA 1217

RESULT 12
 AAS78792/C
 ID AAS78792 standard; cDNA; 3105 BP.
 XX AAS78792;
 AC
 XX
 DR 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #14596
 XX
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX Homo sapiens.
 OS WO200175067-A2.
 PN 11-OCT-2001.
 XX 30-MAR-2001; 2001WO-US08631.
 PF 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX (HYSE-) HYSEQ INC.
 XX Drmanac RT, Liu C, Tang YT;
 XX WPI; 2001-639362/73.
 DR P-PSDB; ABG14605.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX Claim 1; SEQ ID NO 14596; 103bp; English.
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC and responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human

CC diagnostic coding sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.

SO Sequence 3105 BP; 1067 A; 721 C; 658 G; 659 T; 0 other;

Query Match 94.6%; Score 473.8; DB 23; Length 3105;
Best Local Similarity 96.6%; Pred. No. 2.7e-119;
Matches 484; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

```
QY 1 CTTTGGGTTTGTAGACATGAGTCCTGGCCATGCCATGCTGATGATGGCTG 60
D 1717 CTTTGGGTTTGTAGACATGAGTCCTGGCCATGCCATGCTGATGATGGCTG 1658
QY 61 GGTTCCTCTAGAGGATTTTATGCTTTAGGCTCAACATTTAACTTTATCCATCTG 120
D 1657 GGTTCCTCTAGAGGATTTTATGCTTTAGGCTCAACATTTAACTTTATCCATCTG 1598
QY 121 AATTAAATTTTGTATTAAGGTGTAAAGGAAGGATCCAGTTTCAGCTTCTACATAGGCTA 180
D 1597 AATTAAATTTTGTATTAAGGTGTAAAGGAAGGATCCAGTTTCAGCTTCTACATAGGCTA 1538
QY 181 GCCAGTTTCTCAGCACCACATTTATTAATAGGGAATCCTTCCCATTTGCTTGTCT 240
D 1537 GCCAGTTTCTCAGCACCACATTTATTAATAGGGAATCCTTCCCATTTGCTTGTCT 1478
QY 241 CAGGTTTGTCAAAAGATCAGATAGTGTAGATATCGGCAATTTTTCAGAGGCTCTGCTC 300
D 1477 CAGGTTTGTCAAAAGATCAGATAGTGTAGATATCGGCAATTTTTCAGAGGCTCTGCTC 1418
QY 301 TGTTCCTCTAGGATATCTCTGTTTGGTACAGTACATGATGTTTGGTACTGTAG 360
D 1417 TGTTCCTCTAGGATATCTCTGTTTGGTACAGTACATGATGTTTGGTACTGTAG 1358
QY 361 CCTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 420
D 1357 CCTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 1298
QY 421 GGATTTGACTTGGCAGATGTGGGCTCTTTTGGTTCACATGATGATTTTAAAGTACTTTT 480
D 1297 GGATTTGACTTGGCAGATGTGGGCTCTTTTGGTTCACATGATGATTTTAAAGTACTTTT 1238
QY 481 CCAATTCGTGTAGGAAGAATGA 501
D 1237 CCAATTCGTGTAGGAAGAATGA 1217
```

RESULT 13

AA579645/C
ID AA579645 standard; cDNA: 3105 BP.

AC AA579645;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #15449.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN MO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;
PI WPI: 2001-639362/73.
DR P-PSDB: ABG15458.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity

PS Claim 1; SEQ ID No 15449; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AA564197-AA594564 represent novel human
CC diagnostic coding sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.

SO Sequence 3105 BP; 1067 A; 721 C; 658 G; 659 T; 0 other;

Query Match 94.6%; Score 473.8; DB 23; Length 3105;
Best Local Similarity 96.6%; Pred. No. 2.7e-119;
Matches 484; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

```
QY 1 CTTTGGGTTTGTAGACATGAGTCCTGGCCATGCCATGCTGATGATGGCTG 60
D 1717 CTTTGGGTTTGTAGACATGAGTCCTGGCCATGCCATGCTGATGATGGCTG 1658
QY 61 GGTTCCTCTAGAGGATTTTATGCTTTAGGCTCAACATTTAACTTTATCCATCTG 120
D 1657 GGTTCCTCTAGAGGATTTTATGCTTTAGGCTCAACATTTAACTTTATCCATCTG 1598
QY 121 AATTAAATTTTGTATTAAGGTGTAAAGGAAGGATCCAGTTTTCAGCTTCTACATAGGCTA 180
D 1597 AATTAAATTTTGTATTAAGGTGTAAAGGAAGGATCCAGTTTTCAGCTTCTACATAGGCTA 1538
QY 181 GCCAGTTTCTCAGCACCACATTTATTAATAGGGAATCCTTCCCATTTGCTTGTCT 240
D 1537 GCCAGTTTCTCAGCACCACATTTATTAATAGGGAATCCTTCCCATTTGCTTGTCT 1478
QY 241 CAGGTTTGTCAAAAGATCAGATAGTGTAGATATCGGCAATTTTTCAGAGGCTCTGCTC 300
D 1477 CAGGTTTGTCAAAAGATCAGATAGTGTAGATATCGGCAATTTTTCAGAGGCTCTGCTC 1418
QY 301 TGTTCCTCTAGGATATCTCTGTTTGGTACAGTACATGATGTTTGGTACTGTAG 360
D 1417 TGTTCCTCTAGGATATCTCTGTTTGGTACAGTACATGATGTTTGGTACTGTAG 1358
QY 361 CCTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 420
D 1357 CCTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 1298
QY 421 GGATTTGACTTGGCAGATGTGGGCTCTTTTGGTTCACATGATGATTTTAAAGTACTTTT 480
D 1297 GGATTTGACTTGGCAGATGTGGGCTCTTTTGGTTCACATGATGATTTTAAAGTACTTTT 1238
```

QY 481 CCAATTCGTGGAAGAAGTAA 501
|||||
Db 1237 CCAATTCGTGGAAGAAGTCA 1217

RESULT 14
AAS76858/c
ID AAS76858 standard; cDNA; 3239 BP.
XX
XX AAS76858;
XX
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #12662.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PA
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
DR P-PSDB; ABG12671.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 1; SEQ ID NO 12662; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
CC
XX
XX Sequence 3239 BP; 1091 A; 779 C; 686 G; 683 T; 0 other;

Query Match 94.6%; Score 473.8; DB 23; Length 3239;
Best Local Similarity 96.6%; Pred. No. 2.7e-119;
Matches 484; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 1 CTTTGGTGTGTAGACATGAAGTCCTTCCCATGCTTATGTCGAATGATTCGCTG 60
|||||

Db 1717 CTTTGGTGTGTTCACATGAAGTCCTTCCCATGCTTATGTCGAATGATTCGCTA 1658
QY 61 GGTTCCTCTAGGAGTTTATAGTGTTCATCAATTTAAGTCCTTAATCCATCTTG 120
|||||
Db 1657 GGTTCCTCTAGGAGTTTATAGTGTTCATCAATTTAAGTCCTTAATCCATCTTG 1598
|||||
QY 121 AATTAATTTTGTATAGGTGTAAAGGAGATCCAGTTTCAGCTTCTACATAGGCTA 180
|||||
Db 1597 AATTAATTTTGTATAGGTGTAAAGGAGATCCAGTTTCAGCTTCTACATAGGCTA 1538
|||||
QY 181 GCCAGTTTCTCAGCACCAATTATTAATAGGAATCCCTCCCAATGCTGTTTCT 240
|||||
Db 1537 GCCAGTTTCTCAGCACCAATTATTAATAGGAATCCCTCCCAATGCTGTTTCT 1478
|||||
QY 241 CAGGTTTGCACCAATGATAGTGTATATATGCGCATTAATTTGTAGGGCTCTGTC 300
|||||
Db 1477 CAGGTTTGCACCAATGATAGTGTATATATGCGCATTAATTTGTAGGGCTCTGTC 1418
|||||
QY 301 TGTTCATGTTGTATATCTCTGTTTGTACCATGATGTTGTTTGTACTGTAG 360
|||||
Db 1417 TGTTCATGTTGTATATCTCTGTTTGTACCATGATGTTGTTTGTACTGTAG 1358
|||||
QY 361 CCTGTAGTGTAGTTGAAGTCAGTAGCATGATGCTCCAGCTTTGTCCTTTGGCTTA 420
|||||
Db 1357 CCTGTAGTGTAGTTGAAGTCAGTAGCATGATGCTCCAGCTTTGTCCTTTGGCTTA 1298
|||||
QY 421 GGATTCAGTTGGGATGTCGCTTTTGTGTCCACATGAACCTTAAGTATTTT 480
|||||
Db 1297 AGATTGACTTGGCAAGTGGGCTTTTGTGTCCATGAACCTTAAGTATTTT 1238
|||||

QY 481 CCAATTCGTGGAAGAAGTAA 501
|||||
Db 1237 CCAATTCGTGGAAGAAGTCA 1217

RESULT 15
AAS77062/c
ID AAS77062 standard; cDNA; 4114 BP.
XX
XX AAS77062;
XX
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #12866.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PA
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
DR P-PSDB; ABG12875.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 1; SEQ ID NO 12866; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX

SO Sequence 4114 BP: 1438 A; 951 C; 808 G; 917 T; 0 other;

Query Match: 94.6%; Score 473.8; DB 23; Length 4114;

Best Local Similarity 96.6%; Pred. No. 2.9e-119;

Matches 484; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

```
OY 1 CTTTGGTGTATACATGAAGTCCCTGCGCATATGCTGATGATGCGCTG 60
    ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 2587 CTTTGGTGTATACATGAAGTCCCTGCGCATATGCTGATGATGCGCTG 2528
    ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

OY 61 GGTTCCTCAGAGGATTTTAAAGTCTTACATTTAACTTTAATCATCTTG 120
    ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 2527 GGTTCCTCAGAGGATTTTAAAGTCTTACATTTAACTTTAATCATCTTG 2468
    ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

OY 121 AATTAAATTTTGTATAGGTGAAGAGGATCCAGTTTCAGCTTTCTACATAGGCTA 180
    ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 2467 AATTAAATTTTGTATAGGTGAAGAGGATCCAGTTTCAGCTTTCTACATAGGCTA 2408
    ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

OY 181 GCCAGTTTTCAGCAGCACCATTATTAATAAGGAAATCCTTCCCATGCTGTTTCT 240
    ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 2407 GCCAGTTTTCAGCAGCACCATTATTAATAAGGAAATCCTTCCCATGCTGTTTCT 2348
    ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

OY 241 CAGTTTGTCAAAGATCAGATAGTTGTAGATATGCGGCAATATTTCTAGGGCTCTGTT 300
    ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 2347 CAGTTTGTCAAAGATCAGATAGTTGTAGATATGCGGCAATATTTCTAGGGCTCTGTT 2288
    ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

OY 301 TGTTCATTTGTTGATATCTGTTTGTGTACCACTACCATGTTGTTTGTACTGTAG 360
    ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 2287 TGTTCATTTGTTGATATCTGTTTGTGTACCACTACCATGTTGTTTGTACTGTAG 2228
    ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

OY 361 CCTGTAGTGTAGTTGAAGTCAGGTACATGATGCTCCAGCTTGTCTTTGGCTTA 420
    ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 2227 CCTGTAGTGTAGTTGAAGTCAGGTACATGATGCTCCAGCTTGTCTTTGGCTTA 2168
    ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

OY 421 GGATTTGACTGCGCATGTGGGCTCTTTTGGTTCCACATGAACCTTTAAAGTAGTTT 480
    ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 2167 AGATTTGACTGCGCATGTGGGCTCTTTTGGTTCCACATGAACCTTTAAAGTAGTTT 2108
    ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

OY 481 CCAATTTCTGAGAAAGTAA 501
    ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 2107 CCAATTTCTGAGAAAGTCA 2087
    ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
```

Search completed: July 4, 2003, 15:25:06
Job time : 168.778 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 15:09:32 ; Search time 1095.73 Seconds

(without alignments)
7405.060 Million cell updates/sec

Title: US-10-083-853b-2_COPY_18900_19400

Perfect score: 501
Sequence: Tttctatgactaaagcacc.....gcaccctgttctctttt 501

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

EST: *
1: em_estda: *
2: em_esthum: *
3: em_estin: *
4: em_estmu: *
5: em_estrov: *
6: em_estrpl: *
7: em_estro: *
8: em_hlc: *
9: gb_est1: *
10: gb_est2: *
11: gb_hlc: *
12: gb_est3: *
13: gb_est4: *
14: gb_est5: *
15: em_estfun: *
16: em_estom: *
17: gb_gss: *
18: em_gss_hum: *
19: em_gss_inu: *
20: em_gss_pln: *
21: em_gss_vrt: *
22: em_gss_fun: *
23: em_gss_mam: *
24: em_gss_mus: *
25: em_gss_other: *
26: em_gss_pro: *
27: em_gss_rtd: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47	9.4	1101	17	CNS008WC
2	41.6	8.3	1101	17	AL052719 Drosophila
3	40	8.0	559	10	AL098379 Drosophila
4	39.4	7.9	1101	17	BB646396 BB646396
5	39.2	7.8	274	9	CNS006DR
6	38.8	7.7	497	12	AV223573 AV223573
					BF423025 BF423025
					HC_d11_21

C 7	38.8	7.7	574	17	BH539949
C 8	38.8	7.7	885	13	B1546960
C 9	38.8	7.7	993	17	CNS0236T
C 10	38.6	7.7	414	13	B3936625
C 11	38.6	7.7	641	13	B1585460
C 12	38.6	7.7	836	17	A2682240
C 13	38.6	7.7	883	17	A2690387
C 14	38.6	7.7	921	17	BH150526
C 15	38.2	7.6	425	17	BH18375
C 16	38.2	7.6	840	17	A0027476
C 17	38	7.6	837	17	BH165379
C 18	38	7.6	1110	17	CNS01RXV
C 19	37.8	7.5	709	12	B6385089
C 20	37.6	7.5	288	10	BH181383
C 21	37.6	7.5	1204	17	CNS016E2
C 22	37.4	7.5	493	13	B1946460
C 23	37.4	7.5	559	13	B1574743
C 24	37.4	7.5	624	17	A2415638
C 25	37.4	7.5	698	13	B1584493
C 26	37.4	7.5	1024	17	CNS005BT
C 27	37.2	7.4	600	13	B1988264
C 28	37.2	7.4	898	17	CNS01POH
C 29	37	7.4	636	13	BM184301
C 30	37	7.4	647	17	AG065239
C 31	36.8	7.3	414	17	A0045487
C 32	36.8	7.3	491	10	BH781797
C 33	36.8	7.3	828	17	BH689601
C 34	36.8	7.3	919	17	A2538973
C 35	36.6	7.3	424	17	BH352158
C 36	36.6	7.3	480	17	BH763634
C 37	36.6	7.3	939	17	BH137105
C 38	36.6	7.3	1201	17	CNS002OT
C 39	36.4	7.3	252	10	BB076733
C 40	36.4	7.3	391	17	A2410995
C 41	36.4	7.3	500	9	AU088252
C 42	36.4	7.3	695	17	BH161708
C 43	36.4	7.3	870	17	A2677313
C 44	36.4	7.3	919	17	CNS005RL
C 45	36.4	7.3	996	17	CNS00FHH

ALIGNMENTS

RESULT 1
LOCUS CNS008WC 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR18L14 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL052719.1 GI:4934268
VERSION AL052719

KEYWORDS GSS.
SOURCE Drosophila melanogaster.

ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1101)
Genoscope.
AUTHORS Direct Submission
TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr
JOURNAL - Web : www.genoscope.cns.fr)

COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osogawa and
Aaron Mammosier in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial

ECORI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source

Location/Qualifiers

1..1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR18L14"
/clone.lib="RPCL-98"
/note="end : TET3"

BASE COUNT 145 a 171 c 122 g 340 t 323 others

ORIGIN

Query Match 9.4%; Score 47; DB 17; Length 1101;
Best Local Similarity 28.7%; Pred. No. 0.015;
Matches 99; Conservative 84; Mismatches 162; Indels 0; Gaps 0;

QY 8 AACATAAGCACCACATATTTTCAAGCTAGTATTAATTTTACTGATTCATTC 67
DB 703 AAAAAATAATAATAAATTTTATTAATAAATAAATAATTTTATTTTAA 762
QY 68 GCATATATGTGTATATAATATAGATTTGATTTGCTGCTCCAGTCATTCATATC 127
DB 763 ATTTTYYMMWMTTAAATTTTAAATWTTAAWTTTCTTAATATATTAATWMTT 822
QY 128 CATCCATTTTGGCAGCACAATCTTATTTAGTCATACATAGATATCTTCCACA 187
DB 823 WMMWMTTATTTATATAAAMMMWMTTATTTTAAATWMTTAAATTAATTAACACW 882
QY 188 GAACACAGTTAGAACCAAGTCCTTTGCTGCACAGAACCCACCTTTCCAAATGA 247
DB 883 CMCCCATCNCMTAHMCMYCCHCTCAVHMCMYCMCMCCCMCCCMCCCMCCMCC 942
QY 248 CGCCCTTGCTGACAGCCGACACAGAGAAATCCCATGATTTAGCGATATAC 307
DB 943 CCMCCYMACWATTTCTCMTCCACCCACACACACACACACACACACACACAC 1002
QY 308 CATGTTCAATCTTAACCTAGCCCTGTAAATGTTTGAAGCTTAA 352
DB 1003 CAAATATTTTMTTATATHTTTCACACACACACACACACACACACACAC 1047

RESULT 2
CNS0100X/c 1101 bp DNA linear GSS 26-JUL-1999
LOCUS Drosophila melanogaster genome survey sequence SP6 end of BAC
DEFINITION BAC003604 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.

ACCESSION AL098379.1 GI:5609990
VERSION AL098379
KEYWORDS GSS.

SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1101)
Genoscope.

TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)

COMMENT

determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (BDGP) -
<http://www.edgp.edi.ac.uk> - This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.

FEATURES

Location/Qualifiers

source

1..1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACN03604"
/clone.lib="DrosBAC"
/plasmid="pBelobAC11"
/note="end : SP6"

BASE COUNT 195 a 108 c 131 g 161 t 506 others

ORIGIN

Query Match 8.3%; Score 41.6; DB 17; Length 1101;
Best Local Similarity 18.6%; Pred. No. 0.55; Indels 7; Gaps 1;
Matches 65; Conservative 146; Mismatches 132;

QY 43 ATAATATTTTCTTAATTTCAATTCGCAATTTATGTTATTAATGATTTGATTCG 102
DB 1092 HNMWMTTAMTATAGWTKATMMKATMMKATMMKATMMKATMMKATMMKATMM 1033
QY 103 TGTTCCTCAAGTCATCCAAATTAATTCATTCATTTGGCAGCCAAATCTTTATTTGGTCA 162
DB 1032 YKCHYNNMYNMTCHNTCCMYCSCMCSCMCSCMCSCMCSCMCSCMCSCMCSC 973
QY 163 TACATACATAGATCTCTCTCACAGAACAGTGAAGCAAGTCCTTTGCTGCACA 222
DB 972 TCMCTACATCT 913
QY 223 GAACCCACACCTTTCCCAATGCAGCCCTTTGCTGCAGCCACACAGCAAGAT 282
DB 912 HCTMAGTCANHCNCTMAMMYNNMYNNMYNNMYNNMYNNMYNNMYNNMYNNMY 853
QY 283 CCCATGAATG-----TAGGCACTTATCCACATGTCATTTCAATCTTAACTGACCTGCT 335
DB 852 ATMTGCATATATMTATATACANHTMTATADNMHTMMCTYNNMCMYNNMYNN 793
QY 336 AAATGCTTAGGCTTAACCTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 385
DB 792 MNNHTNNHCHMCHYNNHTNNHTNNHTNNHTNNHTNNHTNNHTNNHTNNHTNN 743

RESULT 3
BB646396/c 559 bp mRNA linear EST 26-OCT-2001
LOCUS BB646396 RIKEN full-length enriched, 10 days neonate cerebellum Mus
DEFINITION musculus cDNA clone B930009J12 5', mRNA sequence.
ACCESSION BB646396
VERSION BB646396.1 GI:16480725
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 559)
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,T., Kono,H., Koude,
M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,
Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,
D., Shibata,K., Shinagawa,A., Shirai,T., Sogabe,Y., Suzuki,H.,
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toyata,T.,
Muramatsu,M. and Hayashizaki,Y.

TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Saitama-shi, Tsukuba-shi, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: <http://genome-gsc.riken.go.jp/>
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,
M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* 10 (10), 1617-1630 (2000)

Wagi, K., Fujikawa, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanishi, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, Y., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

RIKEN Integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. *Genome Res.* 10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)

Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamamata, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. *Mamm. Genome* 12, 673-677 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

e mouse tissues.

FEATURES
source

BASE COUNT
ORIGIN

189 a 84 c 132 g 154 t

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="B930009j12"
/clone_lib="RIKEN full-length enriched, 10 days neonate
cerebellum"
/sex="mixed"
/tissue_type="cerebellum"
/dev_stage="10 days neonate"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5',
GAGGAGAGAGAGATCCACAGACGCTTTTCTTTTTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 100.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGGAGAGAGATTTCGAGCTTATTAATTAATCCCCCCCCCCCC
3']. cDNA was cloned into the XhoI and BamHI sites.
Vector: a modified pBluescript KS(+) after bulk excision
from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
BamHI"
```

Query Match	8.0%;	Score 40;	DB 10;	Length 559;
Best Local Similarity	52.4%;	Pred. No. 1.3;		
Matches 88;	Conservative 0;	Mismatches 80;	Indels 0;	Gaps 0;

QY	14	AAGCACCACATATTTTCAAGCTACATCAATATTTACTTGAAATTCATTCGCATTA	73
Db	394	AAGAACACAGTATCTTTAAGGGGCTGGCCCATTAATTTCATTCATGTCAGGAAAT	335
QY	74	TTGTGTTACATAAATTAGATTTGATTTGCTGCTGCCAAGTGCATTCATATTCATCA	133
Db	334	ATGGGCAACATAAATTGACCTTGGCTGATGCTTCCTCTTCTACTACTTCCTCTACTA	275
QY	134	TTTGGCGCAAAATCCTTATGTCGATACATACATAGATATCTTC	181
Db	274	TTTTTCTCCAGTCTTCCTCTCTGCTGCTCCATATGTCATCAATCGTC	227

RESULT 4
CNS00GDR

LOCUS	CMS006DR	1101 bp	DNA	linear	SSS 03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC: BACR33M19 of RPEC-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION	AF072060				

ACCESSION	AL072060
VERSION	AL072060.1
KEYWORDS	GSS.
SOURCE	Drosophila melanogaster

ORGANISM	REFERENCE
Drosophila melanogaster	1
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota	1
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	1
Phyroidae; Drosophilidae; Drosophila.	1
1 (bases 1 to 1101)	1

AUTHORS
TITLE
JOURNAL

COMMENT

The BDGP is constructing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information please see <http://www.fruitfly.org>. The BDGP *Drosophila melanogaster* BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPc1-98 and was constructed by partial EcoRI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source

```

/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR3IM19"
/clone_lib="RPCI-98"
/note="end : 17"

```

BASE COUNT	291 a	178 c	34 g	216 t	382 others
ORIGIN					
Query Match		7.9%	Score 39.4;	DB 17;	Length 1101;
Best Local Similarity		26.8%;	Pred. NO. 2.4;		
Matches	97;	Conservative	88;	Mismatches	177;
				Indels	0;
				Gaps	0;

QY 12 AAAAGCACCAATATTTTTCAAAGCTAAGTCATAAAATATTACTTGGAATTCATGGCAT 71
||| : | : : || : : || : : | | | | | :
Db 554 AAATMAAMTNTAGCMTKTATMAAMTAAMVTAADGATGTATKTTTAAAAAAAATTTATMM 613

QY 72 TATGTGTACATAAATTAGATTTGATTTGGTGTCTTCCAGTCCATCCATTATCCATC 131

Db 614 TTAATAATMTAGIDATRTAMATTAIRWMMATATTATTTTAAAAADTTGTWTAANAANW 673

Db 674 WATTTATGTCGTGAACACMAATDKTTTTTTMTMAGAAATTTATKTTTTTDACTCTRWATCTKA 733

192 ACAGTTAGAGCAAGTCTTTTCTCTGACAGAACCCACACACCTTTCCCATGACGCC 251

Db 734 DAATTWGGTITMTWKATGMTCKCKWAWTKTANMARMAMAMWATRTMAAATWTDW 793

Db 794 ATAKTTTCTRASMCMCGAATAATGKKKGMTTRACWKGARRAKAAAAMKAAAMAKATKXXX 853

[illegible]

Db 854 WKAADKKATITKMSKCTGKAWKKWMTAMFTKKRATDMPITKAAAKGWAALMKCATAPAAA 913

DB		914 AM 915
OY	372 Aa 373 1 :	
RESULT 5		
AV223573		
LOCUS		
DEFINITION	AV223573 RIKEN full-length enriched, 18 days pregnant, placenta and extra embryonic tissue Mus musculus cDNA clone 3830412F23 , mRNA sequence.	
ACCESSION	AV223573	
VERSION	AV223573.1	GI:6172951
KEYWORDS	EST.	
SOURCE	house mouse. Mus musculus	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. (bases 1 to 274)	
REFERENCE	Kono, H., Akizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T., Fukuda, S., Furukashi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, T., Kikuchi, N., Kojima, Y., Koya, S., Kusakabe, M., Matsuyama, T., Mikita, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ow, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Takahashi, F., Tateo, N., Tomimaga, N., Tsunoda, Y., Watabiki, A., Watanabe, S., Yamamura, T., Yasunishi, A., Yokota, T., Yoshida, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.	
AUTHORS	RIKEN Mouse ESTs (Kono, H., et al. 1999) Unpublished (1999)	
TITLE	Contact: Yoshihide Hayashizaki	
JOURNAL	Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute	
COMMENT	The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Seishiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/ Sasaki, N., Izawa, M., Watabiki, M., Ozawa, K., Tanaka, T., Yoneda, Y., Matsura, S., Carninci, P., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.	
	Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998) Itch, M., Kitsuai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.	
	Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999) Carninci, P. and Hayashizaki, Y.	
	High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)	
	Please visit our web site (http://genome.gsc.riken.go.jp) for further details.	
FEATURES	Location/Qualifiers	
source	1..274 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone_id="3830412F23" /clone_lib="RIKEN full-length enriched, 18 days pregnant, placenta and extra embryonic tissue" /sex="female" /tissue_type="placenta and extra embryonic tissue" /dev_stage="18 days pregnant, adult" /lab_host="DH10B" /note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was	

primed with a primer [5'
GAGAGAGAGAGATCCAGAGCTCTTTTTTTTTTTTTTVN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 50.0. Second strand
cDNA was prepared with the primer adapter of sequence [5'
GAGAGAGAGATTCGAGATTATTAATTAATTAATCCCCCCCCCC 3']. cDNA
was cloned into the XhoI and BamHI sites. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC I"

BASE COUNT	82 a	40 c	50 g	102 t
ORIGIN				
Query Match	7.8%;	Score 39.2;	DB 9;	Length 274;
Best Local Similarity	52.4%;	Pred. No. 1.9;		
Matches	86;	Conservative	0;	Mismatches 78; Indels 0; Gaps 0;
Db	10	CTAAAGCAGACATATTTTCAAGACAAATGCATAAATATTACTTGAAATTCATTCGC	69	
Db	27	CCAAATCCACCATGATGATTAATTAATGCTACACCTGTACATCTTGATATAAGTTAATCTT	86	
Qy	70	ATTATTTGTTACATTAATTAATTAATTTGATTTTGTTGCTTCCCAAGTCATCCAAATTAATCA	129	
Db	87	ATTTTGGATTTTAAATTAATTTGGATGGGGGGCTATGAACGTGTAGAGAAATTAAGTTCTC	146	
Qy	130	TCCATTTTGGAGCCAAATCCTTATTAATTTGGTGCATATACATATCATTA	173	
Db	147	CCAAATATTTCATATATTAGTTTTTTTGTACATAGCTTTCTTGA	190	

```

RESULT 6
BF423025/c
LOCUS
DEFINITION
BF423025
497 bp mRNA linear EST 28-NOV-2000
HC_dil_21B10.SKPL Haemonchus contortus d11 mixed adult from David
Knox Haemonchus contortus cDNA clone HC_dil_21B10 5', mRNA
sequence.
BF423025
BF423025.1 GI:11411014
EST.
Haemonchus contortus.
Haemonchus contortus
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
Trichostrongyloidea; Haemonchidae; Haemonchinae; Haemonchus.
1 (bases 1 to 497)
Blaxter,M.L., Parkinson,J., Whitton,C., Daub,J., Guillian,D., Hall
,N., Quayle,M. and Barrell,B.
Edinburgh University/Sanger Centre Nematode EST Project
Unpublished (2000)
Contact: Blaxter ML
Institute of Cell, Animal and Population Biology
University of Edinburgh
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9
3J7, UK.
Tel: +44 131 650 6760
Fax: +44 131 670 5450
Email: mark.blaxter@ed.ac.uk
The library was prepared by David Knox of the Moredun Institute,
Edinburgh. Sequencing was performed by the Pathogen Sequencing Unit
, Sanger Centre, Cambridge, UK (Neil Hall, Mike Quail & Bart
Barrell). The sequence contained a PolyA tail (trimmed)
PCR Primers
FORWARD: T3
BACKWARD: T7PL
Plate: 21 row: B column: 10
Seq primer: SKPL
High quality sequence stop: 497.
Location/Qualifiers
1..497
/organism="Haemonchus contortus"
/strain="Moredun"
/db_xref="taxon:6289"
/cclone="HC_dil_21B10"
FEATURES
source

```

```

/clone.lib="Haemonchus contortus d11 mixed adult from
David Knox"
/dev_stage="Adult, day 11"
/Note="Vector: Lambda Zap II; Site_1: EcoRI; Site_2: XhoI;
Constructed by David Knox, Moredun Institute, Edinburgh.
Primary library filter was 4x106 pfu."
BASE COUNT      207 a      30 c      59 g      201 t
ORIGIN

Query Match      7.7%; Score 38.8; DB 12; Length 497;
Best Local Similarity 60.4%; Pred. No. 2.9;
Matches 64; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 23 AATATTTCAGAGTCATTAATATTACTTGATTCATTCGCATTATGCTTAC 82
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 205 AAAAAATTTAAATTTTACCTATTAATTTTAAATTTTAAACACATTAATTTTATVA 146
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 83 ATAAATTAGATTGATTGCTGCTCCAGATCCATCCATTAATCC 128
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 145 TTAAACAAATAAAGTTAAATTAATTAATGACATGTCGCGATTATTC 100
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
BH539949      574 bp      DNA      linear      GSS 14-DEC-2001
LOCUS      BOHH179TF BOHH Brassica oleracea genomic clone BOHH179, DNA
DEFINITION      sequence.
ACCESSION      BH539949
VERSION      BH539949.1 GI:17784813
KEYWORDS      GSS.
SOURCE      Brassica oleracea.
ORGANISM      Brassica oleracea.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 574)
TOWH,C.D., Van Aken,S., Uteirback,T. and Fraser,C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished (2001)
Other_GSSS: BOHH179TF
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TP
Class: sheared ends.
FEATURES
    source      Location/Qualifiers
                1..574
                /organism="Brassica oleracea"
                /strain="TO100DH3"
                /db_xref="taxon:3712"
                /clone="BOHH179"
                /clone.lib="BOHH"
                /note="Vector: PHOS1; Site_1: BstXI; 2-3 kb sheared
                genomic DNA inserted into PHOS1 using BstXI linkers"
BASE COUNT      215 a      62 c      79 g      218 t
ORIGIN

Query Match      7.7%; Score 38.8; DB 17; Length 574;
Best Local Similarity 60.4%; Pred. No. 3;
Matches 64; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 23 AATATTTCAGAGTCATTAATATTACTTGATTCATTCGCATTATGCTTAC 82
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 107 AATGATTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 48
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 83 ATAAATTAGATTGATTGCTGCTCCAGATCCATCCATTAATCC 128
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 47 AAAAAATTGATTGATTCGCGTATCCAAACTGATCCATAACCC 2
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 8
BI546960/c      885 bp      mRNA      linear      EST 05-SEP-2001
LOCUS      603190175F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5261562 5',
DEFINITION      mRNA sequence.
ACCESSION      BI546960
VERSION      BI546960.1 GI:15434272
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 885)
NIH-MGC http://mhc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgarbs@mail.nih.gov
Tissue Procurement: Miklos Palcovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1659 row: e column: 19
High quality sequence stop: 695.
FEATURES
    source      Location/Qualifiers
                1..885
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:5261562"
                /clone.lib="NIH_MGC_95"
                /tissue_type="hippocampus"
                /lab_host="DH10B"
                /note="Organ: brain; Vector: pBluescriptR (modified
                pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
                ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTNN-3',
                size-selected for average insert size 2.5 kb and
                normalized to R0F 5. This is a primary library enriched
                for full-length clones and constructed using the
                cap-trapper method (Carninci, in preparation). Library
                constructed by M. Brownstein (NIH/NHGRI, National
                Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      254 a      157 c      206 g      268 t
ORIGIN

Query Match      7.7%; Score 38.8; DB 13; Length 885;
Best Local Similarity 53.2%; Pred. No. 3.3;
Matches 82; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 48 TATTACTTGATTAATTCGATTTGTTACATTAATTAATGATTGCTGCTTC 107
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 589 TATTATGTTGATCCAAATGCGATTTTAATGCTTTTATACGTAATGATACACTAC 530
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 108 TCCAGATCCATCAATTAATTCATTCATTTGGAGCCAAATCTTATTTGGTCATCAT 167
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 529 ATCAAGTCAATTAATTTGACTTTCTGACAGAGCTTTCTCGTGTGAGCATAGAA 470
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 168 ACATGAGTATCTTTCACAGACACAGCTTAGAA 201
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 469 ACAAGAGTTTATGAGTCAAGTGCACAGTAGAA 436
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
CNS02J6T      993 bp      DNA      linear      GSS 13-MAY-2000
LOCUS      Tetradon nigroviridis genome survey sequence 17 end of clone
DEFINITION      143A01 of library G from Tetradon nigroviridis, genomic survey
sequence.

```

ACCESSION AL199838
 VERSION AL199838.1 GI:7837989
 KEYWORDS GSS: genome survey sequence.
 SOURCE Tetraodon nigroviridis.
 ORGANISM Tetraodon nigroviridis.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Perciformes; Tetraodontiformes; Tetraodontidae; Tetraodon.
 REFERENCE 1 (bases 1 to 993)
 AUTHORS Roest-Crollius H., Jallion O., Dasilva C., Bouneau L., Fisher C., Bernot A., Fizames C., Wincker P., Brothier P., Quetier F., Saurin W. and Weissenbach J.
 TITLE Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 993)
 AUTHORS Roest-Crollius H., Jallion O., Dasilva C., Bouneau L., Fisher C., Bouneau L., Billault A., Quetier F., Saurin W., Bernot A. and Weissenbach J.
 TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 993)
 AUTHORS Genome.
 TITLE Direct Submission
 JOURNAL Submitted (12-Apr-2000)
 COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/tetraodon>.
 FEATURES
 source
 1..993
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone="143A01"
 /clone_lib="G"
 /note="Genoscope sequence ID : C0AG143A01LP1-end : T7"
 BASE COUNT 215 a 197 c 194 g 386 t 1 others
 ORIGIN
 Query Match 7.7%; Score 38.8; DB 17; Length 993;
 Best Local Similarity 56.2%; Pred.No.3.4; Indels 0; Gaps 0;
 Matches 73; Conservative 0; Mismatches 57;
 ORGANISM
 4 TATGAACTAAAGACACCAATATTTTCAAGCTAAGCTAATATTTTACTTGATCA 63
 DB 418 TACAAGATACATTTATACAAATAATGTGACACCAAGTCAACATATTCGTGATCA 359
 QY 64 ATTGCATTATTTGTTTACATTAATTTGATTTGTTGTTCTCCAAATCCATCAAT 123
 DB 358 AAGAAATAATTTGCTCCCTAAATTAAGGAGTGAATATTTATTTTCAAGCCATCAAA 299
 QY 124 TATCCATCA 133
 DB 298 TAAAAAGCCA 289
 RESULT 10
 B393625 414 bp mRNA linear EST 08-MAR-2002
 LOCUS B393625 Dictyostelium discoidium cDNA library, SF Dictyostelium
 DEFINITION discoidium cDNA clone dds32c06 5', mRNA sequence.
 ACCESSION B393625
 VERSION B393625.1 GI:19304711
 KEYWORDS EST.
 SOURCE Dictyostelium discoidium.
 ORGANISM Dictyostelium discoidium.
 Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
 REFERENCE 1 (bases 1 to 414)
 AUTHORS Urushihara H., Tanaka Y., Kohara Y. and Shin-I T.
 TITLE Full length cDNA of Dictyostelium discoidium at the slug stage
 JOURNAL Unpublished (2002)

COMMENT Contact: Tadao Shin-I
 Center for Genetic Resource Information
 National Institute of Genetics
 111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshin@genes.nig.ac.jp.
 FEATURES
 source
 1..414
 /organism="Dictyostelium discoidium"
 /strain="AX4"
 /db_xref="taxon:44689"
 /clone="dds32c06"
 /clone_lib="Dictyostelium discoidium cDNA library, SF"
 /sex="mat A"
 /dev_stage="Slug stage"
 BASE COUNT 188 a 30 c 21 g 160 t 15 others
 ORIGIN
 Query Match 7.7%; Score 38.6; DB 13; Length 414;
 Best Local Similarity 56.3%; Pred.No.3.1;
 Matches 71; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
 ORGANISM
 5 ATGAACTAAAGACACCAATATTTTCAAGCTAAGCTAATATTTTACTTGATCA 64
 DB 179 ATTAACATTAATCTCCAAATAATTTCTTAATCAAAAAAATAATTAATATATATATA 238
 QY 65 TTGCATTATTTGTTTACATTAATTTGATTTGTTGTTCTCCAAATCCATCAAT 124
 DB 239 TTATGTTTGTGTTGTTTAAATTAATTTGTTTGTGTCATCAATTTATTTT 298
 QY 125 ATCAT 130
 DB 299 AAAAT 304
 RESULT 11
 B1585460 641 bp mRNA linear EST 06-SEP-2001
 LOCUS B1585460
 DEFINITION RH25047.5prime RH Drosophila melanogaster normalized Head p1c-1
 Drosophila melanogaster cDNA clone RH25047 5, mRNA sequence.
 ACCESSION B1585460
 VERSION B1585460.1 GI:15476882
 KEYWORDS EST.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 641)
 AUTHORS Stapleton M., Brocksiehn P., Hong L., Tyler D., Berman B., Carlson J., Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R., Gonzalez M., Guarin H., Harris N., Li P., Liao G., Mitra S., Mungall C.J., Nunoo J., Pacleby J., Paragas V., Park S., Phouanavong S., Wan K., Yu C., Lewis S.E., Ceiniker S. and Rubin G.M.
 TITLE BDGP/HMT RH Drosophila EST Project
 JOURNAL Unpublished (2001)
 COMMENT Contact: Stapleton, M.
 BDGP
 Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: <http://www.fruitfly.org/EST/estefruitfly.berkeley.edu>
 hit genomic AB003431: arm:x [4198619,4510082]
 estimated-cyto-4C1-4C8: 08/18/2001
 Plate: RH.250 row: D column: 11
 High quality sequence stop: 516.
 FEATURES
 source
 1..641
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="RH25047"

	Query Match	7.7%	Score 38.6	DB 13	Length 641
	Best Local Similarity	52.2%	Pred. No. 3.5		
	Matches	83	Conservative	0	Mismatches 76
					Indels 0
					Gaps 0
QY	12	AAAAGCACACAAATATTTTCAAAGCTAGTCATAAATATTACTGAATTCAAATTCGCAT	71		
Db	370	AAATTCATTACCAACTTTCAAAGATTAGTGGCAAAATTTATAGATTTCCTTACAACT	429		
QY	72	TATGTGTACATAAATTAGATTGATTTTCGTTTCGCAAGTCATCAATTAATTCATC	131		
Db	430	TATGTATTAGCATAAACTTGGTCAAAATTCGGTTCGTCAAAGCCCNCTTAATTNTCTTC	489		
QY	132	CATTGGAGCCAAATCCTCTTATTTGGCATACATACATA	170		
Db	490	GCATTTTACGGTAATTCATTTGCATAGCTGGCCAAATTTA	528		

	RESULT	12		
	AZ682240			
	LOCUS	836 bp	DNA	linear GSS 14-DEC-2000
	DEFINITION	ENTLNT7OTR Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic , DNA sequence.		
	ACCESSION	AZ682240		
	VERSION	AZ682240.1 GI:11819386		
	KEYWORDS	GSS.		
SOURCE	Entamoeba histolytica . Eukaryota; Eumetazoa; Excavata; Entameoba histolytica. I (bases 1 to 836) Loftus,B., Van Aken,S. and Fraser,C. Determination of clone end sequences from Entamoeba histolytica HM1:IMSS sheared DNA library Unpublished (2000)			
JOURNAL COMMENT	Contact: Brendan J Loftus Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, Tel.: 301 838 0208 Fax: 301 838 3543 Email: bjl@fususetigr.org			
	Clones are derived from the Entamoeba histolytica HM1:IMSS sheared DNA library			
	Seq primer: M13--Reverse			
	Class: shotgun			
	High quality sequence start:.15 High quality sequence stop: 819.			
FEATURES	Location/Qualifiers			
source	I..836			

`/strain="HNL:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entomoea
/notes="Entomoea histolytica Sheared DNA"
institute="Genomic Research (PIGR), Rockville, MD.
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entomoea histolytica: a
77-450). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v +1 method used for
the library construction is described in detail in Smith.`

Query Match	7.7%	Score 38.6;	DB 17;	Length 836;
Best Local Similarity	49.3%;	Pred. No. 3.7;		
Matches 101; Conservative	0;	Mismatches 104;	Indels 0;	Gaps 0

QY	DB	QY	DB
1	AAAGCACCACATATTTTCAAGCAGACATATAATTTCTGGATTCATTCGGAT	72	
49	ACACCAATAGAAATTCACGATATATTTAAATTAAGCATATTTAATATATATTTATT	108	
73	ATTGCTGCATTAATTAAGATTGATTTGATTTGCTTCCAGCTCATCCATTAATCATCC	132	
109	AATATATCATTTCTTTATATATTTTGTGTTGGTTATTAATCTAATTTATGTTTGAATGA	168	
133	ATTGGCAGCCAAATCTCTTATTGGTCATACATACATGAGTATCTTTCACAGAGACA	192	
169	TTTATCATTAATTAACACATATTTTATTCATCTCTAAATATTTTATTCATTAATTA	228	
193	CAGTTAGAGCAAGTCCTTTGCT	217	
229	ATTTCAATTAATATTTCTCTTTAA	253	

FEATURES	source
RESULT 13	
A2690387/C	
LOCUS	
DEFINITION	883 bp DNA linear GSS 14-DEC-2000
ACCESSION	A2690387
VERSION	ENTR4441F
KEYWORDS	Entamoeba histolytica sheared DNA genomic, DNA sequence.
SOURCE	A2690387
ORGANISM	A2690387.1 GI:11827533 GSS.
REFERENCE	Entamoeba histolytica.
AUTHORS	Entamoeba histolytica.
TITLE	Eukaryota; Entamoebidae; Entamoeba.
JOURNAL	1 (bases 1 to 883)
COMMENT	Loftus,B., Van Aken,S. and Fraser,C.
	Determination of clone end sequences from Entamoeba histolytica
	HMI:IMSS sheared DNA library
	unpublished (2000)
	Contact: Brendan J Loftus
	Department of Eukaryotic Genomics
	The Institute for Genomic Research
	9712 Medical Center Dr., Rockville, MD 20850, USA
	Tel: 301 838 0208
	Fax: 301 838 3543
	Email: b.loftus@igrr.org
	Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
	DNA library
	Seq primer: M13-Forward
	Class: shotgun
	High quality sequence start: 103
	High quality sequence stop: 860.
	1. 883 Location/Qualifiers

/db.xref="taxon:5159"
/clone.lib="Entamoeba histolytica sheared DNA"
/note="vector: pROST. Site 1. Bst I; (constructed at The
Institute for Genome Research (iIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The λ + i method used for
the library construction is described in detail in Smith
H.O. and Venter, J.C. (Making small insert libraries for

whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).

BASE COUNT 369 a 89 c 101 g 324 t

Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).

BASE COUNT 349 a 98 c 82 g 392 t

Query Match 7.7%; Score 38.6; DB 17; Length 883;
Best Local Similarity 49.3%; Pred. No. 3.8; Indels 0; Gaps 0;
Matches 101; Conservative 0; Mismatches 104;

Query Match 7.7%; Score 38.6; DB 17; Length 921;
Best Local Similarity 49.3%; Pred. No. 3.8; Indels 0; Gaps 0;
Matches 101; Conservative 0; Mismatches 104;

QY 13 AAGCACCACATATTTTCAAGCTAAGCTAATTAATTTACTTGAATTCATTCGACAT 72
DB 840 ACACCAATAGATTTACAGTATATTAATAAAGATTTATTAATTAATTTATTTAT 781
QY 73 ATTTGTTACATTAATTAATTAATTTGATTTGCTTCCAGTCATCCAAATTAATCC 132
DB 780 AATATATCATTTCTTTTATTTTGTGTTGTATTAATGATTTATGTTTGAATGA 721
QY 133 ATTTGGCAGCCAAATCCCTTATTTGTCATATACATGATGATCTTCTCACAAGACA 192
DB 720 TTTATGATTAATTAACATATTTTATTCATTTTAAATTAATTAATTAATTAATTA 661
QY 193 CAGTTAGAACGACGCTTTTGTCT 217
DB 660 ATTCAAAATTAATTTCTCTTTAT 636

QY 13 AAGCACCACATATTTTCAAGCTAAGCTAATTAATTTACTTGAATTCATTCGACAT 72
DB 84 ACACCAATAGATTTACAGTATATTAATAAAGATTTATTAATTAATTTATTTAT 143
QY 73 ATTTGTTACATTAATTAATTAATTTGATTTGCTTCCAGTCATCCAAATTAATCC 132
DB 144 AATATATCATTTCTTTTATTTTGTGTTGTATTAATGATTTATGTTTGAATGA 203
QY 133 ATTTGGCAGCCAAATCCCTTATTTGTCATATACATGATGATCTTCTCACAAGACA 192
DB 204 TTTATGATTAATTAACATATTTTATTCATTTTAAATTAATTAATTAATTAATTA 263
QY 193 CAGTTAGAACGACGCTTTTGTCT 217
DB 264 ATTCAAAATTAATTTCTCTTTAT 288

RESULT 14
BH150526 921 bp DNA linear GSS 28-AUG-2001
LOCUS ENTPIV04TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION genomic, DNA sequence.

RESULT 15
BH318375 425 bp DNA linear GSS 03-DEC-2001
LOCUS CH230-40L10.TJ CHORI-230 Segment 1 Rattus norvegicus genomic clone
DEFINITION CH230-40L10, DNA sequence.

ACCESSION BH150526
VERSION BH150526.1 GI:15322140
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
REFERENCE 1 (bases 1 to 921)
AUTHORS Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HMI:IMSS sheared DNA library (2001)
COMMENT Unpublished (2001)
CONTACT: Brendan J Loftus
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjoftus@tifgr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 18
High quality sequence stop: 690.

REFERENCE 1 (bases 1 to 425)
AUTHORS Zhao, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K., Shvartsbeyn, A., Gebregeorgis, E., Overton, L., Russell, D., Chen, D., Riggs, F., de Jong, P. and Fraser, C.M.
TITLE Rat BAC End Sequences from Library CHORI-230 EcoRI segment
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: CH230-40L10.TV
CONTACT: Shaying Zhao
Department of Eukaryotic Genomics
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tifgr.org
Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rat230.htm). For BAC library availability, please contact Pieter de Jong (pjejong@small.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/or_ering_information.htm). BAC end
pages: http://www.tifgr.org/lcd/bac_ends/rat/bac_end_intro.html
Plate: 40 row: L column: 10
Seq primer: SP6
Class: BAC ends.

FEATURES
source

FEATURES
source

/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHOST1; Site 1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G. and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a light size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome

Location/Qualifiers
1. 425
/organism="Rattus norvegicus"
/strain="BN/SSHsd/MCM"
/db_xref="taxon:10116"
/clone="CH230-40L10"
/sex="Female"
/cell_type="Brain"
/note="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI; CHORI-230 Rat (BN/SSHsd/MCM) BAC library produced by

[illegible]

QY	241	CAATGCACGGCCCTTGTCGTCACCGGCGACACAGAAAGGAAATCCATGATGTTAGGCGAG	300
Db	19140	CAATGCACGGCCCTTGTCGTCACCGGCGACACAGAAAGGAAATCCATGATGTTAGGCGAG	19199
QY	301	TTATACCACTGTTCAATCTTAACTTAACCTTGTAATGTCCTTAGGCTTAAACTGAAC	360
Db	19200	TTATACCACTGTTCAATCTTAACTTAACCTTGTAATGTCCTTAGGCTTAAACTGAAC	19259
QY	361	TGCTTAAGGAGAAAGAAAAGAGCCCTTGGGACACAGGTCGCAAAATCAAGGACAGT	420
Db	19260	TGCTTAAGGAGAAAGAAAAGAGCCCTTGGGACACAGGTCGCAAAATCAAGGACAGT	19319
QY	421	GTCAGTGAATGCATCTAAGAAAAATTCAGAAACAAGTGTGACCATGATGATCCACTAAC	480
Db	19320	GTCAGTGAATGCATCTAAGAAAAATTCAGAAACAAGTGTGACCATGATGATCCACTAAC	19379
QY	481	CGCATCCCTGTTGTCCTTTT 501	
Db	19380	CGCATCCCTGTTGTCCTTTT 19400	

RESULT 2

```

Sequence 118, Application US/10239676
: Publication No. US20030082609A1
:
GENERAL INFORMATION:
:
APPLICANT: OLEK, Alexander
:
APPLICANT: PIEPENBROCK, Christian
:
APPLICANT: BERLIN, Kurt
:
TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
:
FILE REFERENCE: 5013.1003
:
CURRENT APPLICATION NUMBER: US/10/239,676
:
CURRENT FILING DATE: 2002-09-24
:
PRIORITY APPLICATION NUMBER: PCT/EP01/03968
:
DE 10019058.8
:
DE 10019173.8
:
DE 10032529.7
:
DE 10043826.1
:
PRIORITY FILING DATE: 2001-04-06
:
2000-04-06
:
2000-04-07
:
2000-06-30
:
2000-09-01
:
NUMBER OF SEQ ID NOS: 228
:
SEQ ID NO 118
:
LENGTH: 11036
:
TYPE: DNA
:
ORGANISM: Artificial Sequence
:
FEATURE:
:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
:
NAME/KEY: unsure
:
LOCATION: (9056, 9079, 9081, 9094, 9108..)
:
US-10-239-676-118

```

[illegible]

RESULT 3
ITS-09-03

Sequence 83, Application US/09939980
Patent No. US20020082234A1
GENERAL INFORMATION:

1 APPLICANT: Black, Michael
 2 Burnham, Martin
 3 Hodgson, John
 4 Knowles, David
 5 Itonetto, Michael
 6 Nicholas, Richard
 7 Pratt, Julie
 8 Reichard, Richard
 9 Rosenberg, Martin
 10 Ward, Judith
 11
 12 TITLE OF INVENTION: No. US20020082234A1el Prokaryotic Polynucleotides
 13 Polypeptides and Their Uses
 14
 15 NUMBER OF SEQUENCES: 534
 16
 17 CORRESPONDENCE ADDRESS:
 18 ADDRESSEE: Smithkline Beecham Corporation
 19 STREET: 709 Swedeland Road
 20 CITY: King of Prussia
 21 STATE: PA

```

1  COMPUTER READABLE FORM:
2
3  MEDIUM TYPE: Diskette
4
5  COMPUTER: IBM Compatible
6
7  OPERATING SYSTEM: DOS
8
9  SOFTWARE: FastSEI for Windows Version 2.0
10
11 CURRENT APPLICATION DATA:
12   APPLICATION NUMBER: US/09/939,980
13   FILING DATE: 27-Aug-2001
14   CLASSIFICATION: <Unknown>
15
16 PRIOR APPLICATION DATA:
17   APPLICATION NUMBER: 08/936,165
18   FILING DATE: <Unknown>
19
20 ATTORNEY/AGENT INFORMATION:
21   NAME: Gimmil, Edward R
22   REGISTRATION NUMBER: 38,891
23   REFERENCE/DOCKET NUMBER: P50549
24
25 TELECOMMUNICATION INFORMATION:
26   TELEPHONE: 610-270-4478
27   TELEFAX: 610-270-5090
28   TELEX: <Unknown>
29
30 INFORMATION FOR SEQ ID NO: 83:
31   SEQUENCE CHARACTERISTICS:
32     LENGTH: 924 base pairs
33     type: nucleic acid
34     STRANDEDNESS: single
35     TOPOLOGY: linear
36   MOLECULE TYPE: Genomic DNA
37   SEQUENCE DESCRIPTION: SEQ ID NO: 83:
38
39   US-09-939-980-83

```

	Query Match	7.2%	Score 36.2;	DB 10;	Length 924;	
	Best Local Similarity	51.7%;	Pred. No. 4;	Mismatches	62; Conservative	0; Indels
						Gaps 0;
QY	25 TAATTTCAAGCGTAGTCATAAATTACTGTGTAATCCATATTCGCATATTGGTGTACAT	84				
Dd	536 TAAITTGGAICTTTTATACACCATTAACTAGACGTATNNNNGCATTTTNATATGATAT	477				
QY	85 AAATAGATTGATTTGGTGTTCACAAGTCACATTCATTCACATTTTGGCAGCCA	144				
Dd	476 CAAATGGCGCTNNATTTCTTGTCCTTNNNNATGCATCTTATAGANNNTINGGAGAGCA	417				

RESULT 4

US-10-091-504-2307/c
Sequence 2307, Application US/10091504
Publication No. US20030055908A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC007C1
CURRENT APPLICATION NUMBER: US/10/091,504
CURRENT FILING DATE: 2002-03-07

[illegible]

Query Match	7.2%	Score	36.2;	DB	9;	Length	993;
Best Local Similarity	53.1%	Pred.	No. 4.1;				
Matches	77;	Conservative	0;	Mismatches	68;	Indels	0;
				Gaps	0;		

RESULT 6
US-09-764-869-2307/c
Sequence 2307, Application US/09764869
Patent No. US20020061521A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OR INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC007
CURRENT APPLICATION NUMBER: US/09/764,869
CURRENT FILING DATE: 2001-01-17
Prior application data removed - refer to PAlM or file wrapper

Query Match	7.28;	Score 36.2;	DB 10;	Length 993;
Best Local Similarity	53.18;	Pred. No. 4.1;		
Matches 77; Conservative	0;	Mismatches 68;	Indels 0;	Gaps 0;

```

RESULT 7
US-09-764-869-2308/C
: Sequence 2308, Application US/09764869
: Patent No US20020061521A1
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.,
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PC007
: CURRENT APPLICATION NUMBER: US/09/764,869
: CURRENT FILING DATE: 2001-01-17
: Prior application data removed - refer to PALM or file wrap
: NUMBER OF SEQ ID NOS: 2442
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 2308
: LENGTH: 993
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-764-869-2308

```

RESULT 8
US-09-938-842A-4432/C
Sequence 4432, Application US/09938842A
Patent No. US20020160378A1
GENERAL INFORMATION:
APPLICANT: Harper, Jeff
APPLICANT: Kieps, Joel
APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS,
TRANSGENIC PLANTS CONTAINING
FILE REFERENCE: SCRIPI300-3

APPLICANT: kosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PJ207C1
CURRENT APPLICATION NUMBER: US/10/103 313

```

; NUMBER OF SEQ ID NOS: 653
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 283
; LENGTH: 1954
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: (1598)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc-feature
; LOCATION: (1599)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc-feature
; LOCATION: (1907)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc-feature
; LOCATION: (1915)
; OTHER INFORMATION: n equals a,t,g, or c
;
US-10-103-313-282

```

Query Match	6.7%	Score 33.6;	DB 9;	Length 1954;
Best Local Similarity	63.8%;	Pred. No. 33;		
Matches 51; Conservative	0;	Mismatches 29;	Indels 0;	Gaps 0;

0y 89 TAGATTTGATTTTCGTTTCACAAATCCAAATATCCATTCATTTGGCAGCCAAATC 148
 675 TAGCTGCCATTTCACTGGTCTTACCCCAAGGATTATCCATTCATTTGTGACCAATTC 616
 Db

Qy	149	CTCTATTGGTGATACATA	168
Db	615	CTCTCATAGTTCATTACA	596

RESULT 15
US-10-239

```

Sequence 45 Application US/10239676
Publication No. US20030082609A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
FILE REFERENCE: 5013.1003
CURRENT APPLICATION NUMBER: US/10/239.676
CURRENT FILING DATE: 2002-09-24
PRIORITY APPLICATION NUMBER: PCT/EP01/03968
DE 10019058.8
DE 10019173.8
DE 10032529.7
DE 10043826.1
PRIORITY FILING DATE: 2001-04-06
2000-04-06
2000-04-07
2000-06-30
2000-09-01
NUMBER OF SEQ ID NOS: 228
SEQ ID NO 45
LENGTH: 6544
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-45

```

Query Match	6.6%;	Score 33.2;	DB 9;	Length 6544;
Best Local Similarity	48.9%;	Pred. NO. 82;		
Matches 89;	Conservative 0;	Mismatches 93;	Indels 0;	Gaps 0;

[illegible]

Oy	62	CAATTGGCATTTATGTGTACATTAATAGATTGATTCGTGTTCCAGATCCATGCA	121
Oy		1304	TAACTCTATATACCATTTTACTTTTAAATTAATTACTCAAAATTCACAAATTCACAT
Oy	122	ATTATTCATCCATTTGGCAGCCAAATCCCTATATGGCGCATACATCATGATATCTTC	181
Oy	1244	ATTAACATATATCATCTTCTCTATAAAAAAAATAATTAATTAACCATTAATAATCAAAATATCTTC	1185
Oy	182	TC	183
Oy	1184	TC	1183

Search completed: July 4, 2003, 19:36:13
Job time : 114.075 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OW nucleic - nucleic search, using sw model

Run on: July 4, 2003, 15:16:37; Search time 35.0233 Seconds

(without alignments)
4386.938 Million cell updates/sec

Database 8 Issued: Patentes NANO
Perfect Score 501 US-10-083-853b-2 COPY_18900_19400
Sequence: 1 tctatgaactaaagcacc.....gcattccgtgttcttttt 501

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

1: /cgn2_6/prodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/1/ina/5A_COMB.seq:*
4: /cgn2_6/prodata/1/ina/5B_COMB.seq:*
5: /cgn2_6/prodata/1/ina/PCRTUS_COMB.seq:*
6: /cgn2_6/prodata/1/ina/backfilest1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	36.2	7.2	924	US-08-936-165A-83	Sequence 83, Appl
C 2	35.8	7.1	3030	US-09-693-146-3	Sequence 3, Appl
C 3	33	6.6	6407	US-07-977-630-1	Sequence 1, Appl
C 4	33	6.6	6407	US-07-977-630-2	Sequence 2, Appl
C 5	33	6.6	7218	US-08-232-463-14	Sequence 14, Appl
C 6	32.8	6.5	5558	US-08-961-527-103	Sequence 103, Appl
C 7	32.6	6.5	37950	US-09-338-907-183	Sequence 183, Appl
C 8	32.6	6.5	37950	US-09-218-207-183	Sequence 183, Appl
C 9	32.4	6.4	811	US-09-662-250A-15	Sequence 15, Appl
C 10	32.2	6.4	188	US-08-330-108-2	Sequence 2, Appl
C 11	32.2	6.4	188	PCT-US92-10087-2	Sequence 2, Appl
C 12	32.2	6.4	731	US-08-451-105A-2	Sequence 2, Appl
C 13	32.2	6.4	1223	US-08-463-081B-3	Sequence 3, Appl
C 14	32.2	6.4	1223	US-08-461-379A-3	Sequence 3, Appl
C 15	32.2	6.4	1223	US-08-462-390B-3	Sequence 3, Appl
C 16	32.2	6.4	1223	US-08-463-074B-3	Sequence 3, Appl
C 17	32.2	6.4	1223	US-08-465-585C-3	Sequence 3, Appl
C 18	32.2	6.4	1223	US-08-652-446-3	Sequence 3, Appl
C 19	32.2	6.4	1278	US-09-134-001C-2817	Sequence 2817, Appl
C 20	31.8	6.3	4915	US-09-173-053-7	Sequence 7, Appl
C 21	31.8	6.3	5676	US-08-663-998-3	Sequence 4, Appl
C 22	31.8	6.3	5676	US-08-663-998-4	Sequence 4, Appl
C 23	31.8	6.3	5845	US-09-173-053-1	Sequence 1, Appl
C 24	31.8	6.3	5899	US-09-173-053-2	Sequence 2, Appl
C 25	31.8	6.3	5900	US-08-663-998-1	Sequence 1, Appl
C 26	31.8	6.3	5952	US-08-663-998-2	Sequence 2, Appl
C 27	31.6	6.3	353	US-09-366-887A-21	Sequence 21, Appl

28	31.6	6.3	605	US-09-366-887A-26
C 29	31.6	6.3	44453	US-09-146-053-5
C 30	31.2	6.2	1695	US-09-134-001C-2302
C 31	31.2	6.2	3301	US-08-447-430A-42
C 32	31.2	6.2	3423	US-08-447-430A-40
C 33	31.2	6.2	3423	US-09-485-737B-86
C 34	31.2	6.2	3474	US-08-447-430A-41
C 35	31.2	6.2	3474	US-08-318-837-10
C 36	31.2	6.2	4009	US-08-500-860A-2
C 37	31	6.2	282	US-09-134-001C-2050
C 38	31	6.2	918	US-09-134-001C-355
C 39	31	6.2	3892	US-08-569-214-1
C 40	31	6.2	3892	US-08-937-236-1
C 41	30.8	6.1	3474	US-08-122-458B-8
C 42	30.4	6.1	713	US-08-385-335A-10
C 43	30.4	6.1	6235	US-08-659-206A-4
C 44	30.4	6.1	7287	US-08-659-206A-1
C 45	30.2	6.0	854	US-08-998-416-534

ALIGNMENTS

RESULT 1
US-08-936-165A-83/c
Sequence 83, Application US/08936165A
Patent No. 6348582
GENERAL INFORMATION:
APPLICANT: Black, Michael
APPLICANT: Burnham, Martin
APPLICANT: Hodgson, John
APPLICANT: Knowles, David
APPLICANT: Lonetto, Michael
APPLICANT: Nicholas, Richard
APPLICANT: Pirati, Julie
APPLICANT: Reichard, Richard
APPLICANT: Rosenber, Martin
APPLICANT: Ward, Judith
TITLE OF INVENTION: No. 6348582el Prokaryotic Polynucleotides,
TITLE OF INVENTION: Polypeptides and Their Uses
NUMBER OF SEQUENCES: 534
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/936,165A
FILING DATE: 24-SEP-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/027,032
FILING DATE: 24-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmil, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P50549
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 83:
SEQUENCE CHARACTERISTICS:
LENGTH: 924 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

Sequence 26, Appl
Sequence 5, Appl
Sequence 2302, Ap
Sequence 42, Appl
Sequence 40, Appl
Sequence 86, Appl
Sequence 41, Appl
Sequence 10, Appl
Sequence 2, Appl
Sequence 2050, Ap
Sequence 355, Ap
Sequence 1, Appl
Sequence 1, Appl
Sequence 8, Appl
Sequence 10, Appl
Sequence 4, Appl
Sequence 1, Appl
Sequence 534, App

Sequence 183, Application US/09338907

```
; Patent No. 6265546
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Ilya, Chumakov
; APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: PROSTATE CANCER GENE
; FILE REFERENCE: GENSET.18CP1CP
; CURRENT APPLICATION NUMBER: US/09/338,907
; EARLIER FILING DATE: 1999-06-23
; EARLIER APPLICATION NUMBER: 08/996,306
; EARLIER FILING DATE: 1997-12-22
; EARLIER APPLICATION NUMBER: 60/099,658
; EARLIER FILING DATE: 1998-09-09
; EARLIER APPLICATION NUMBER: 09/218,207
; EARLIER FILING DATE: 1998-12-22
; NUMBER OF SEQ ID NOS: 578
; SOFTWARE: Patent.pm
; SEQ ID NO 183
; LENGTH: 37950
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: exon
; LOCATION: 5259..5328
; OTHER INFORMATION: exon2
; FEATURE:
; NAME/KEY: exon
; LOCATION: 12675..12791
; OTHER INFORMATION: exon3
; FEATURE:
; NAME/KEY: exon
; LOCATION: 14621..14710
; OTHER INFORMATION: exon4
; FEATURE:
; NAME/KEY: exon
; LOCATION: 19822..19912
; OTHER INFORMATION: exon5
; FEATURE:
; NAME/KEY: exon
; LOCATION: 21789..21950
; OTHER INFORMATION: exon6
; FEATURE:
; NAME/KEY: exon
; LOCATION: 23387..23510
; OTHER INFORMATION: exon7
; FEATURE:
; NAME/KEY: exon
; LOCATION: 25520..26016
; OTHER INFORMATION: exon8
; US-09-338-907-183

Query Match
Best Local Similarity 57.3%; Score 32.6; DB 4; Length 37950;
Matches 59; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

OY 28 TTTCAGAGCTAGCTAATAATATTACTGGAATTCGATTCGATTTGTTACATAAA 87
Db 5889 TTAAATAATATTAATCTTAATATTTAAGTAAGTAAGTAAGCTTTGTTTAAACATAAA 5948
OY 88 TTGAGATTGATTTGCGTGTCTCCAGATCCATCCATTAATCCAT 130
Db 5949 TTTCGTGTGATTTGAATCTGCGCTTGCAATTTTATTACTT 5991

RESULT 8
US-09-218-207-183
; Sequence 183, Application US/09218207
; Patent No. 6346381
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Ilya, Chumakov
```

```
; APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: Prostate cancer gene
; FILE REFERENCE: GENSET.018CP1
; CURRENT APPLICATION NUMBER: US/09/218,207
; EARLIER FILING DATE: 1998-12-22
; EARLIER APPLICATION NUMBER: 08/996,306
; EARLIER FILING DATE: 1997-12-22
; EARLIER APPLICATION NUMBER: 60/099,658
; EARLIER FILING DATE: 1998-09-09
; NUMBER OF SEQ ID NOS: 578
; SOFTWARE: Patent.pm
; SEQ ID NO 183
; LENGTH: 37950
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: exon
; LOCATION: 5259..5328
; OTHER INFORMATION: exon2
; FEATURE:
; NAME/KEY: exon
; LOCATION: 12675..12791
; OTHER INFORMATION: exon3
; FEATURE:
; NAME/KEY: exon
; LOCATION: 14621..14710
; OTHER INFORMATION: exon4
; FEATURE:
; NAME/KEY: exon
; LOCATION: 19822..19912
; OTHER INFORMATION: exon5
; FEATURE:
; NAME/KEY: exon
; LOCATION: 21789..21950
; OTHER INFORMATION: exon6
; FEATURE:
; NAME/KEY: exon
; LOCATION: 23387..23510
; OTHER INFORMATION: exon7
; FEATURE:
; NAME/KEY: exon
; LOCATION: 25520..26016
; OTHER INFORMATION: exon8
; US-09-218-207-183

Query Match
Best Local Similarity 57.3%; Score 32.6; DB 4; Length 37950;
Matches 59; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

OY 28 TTTCAGAGCTAGCTAATAATATTACTGGAATTCGATTCGATTTGTTACATAAA 87
Db 5889 TTAAATAATATTAATCTTAATATTTAAGTAAGTAAGTAAGCTTTGTTTAAACATAAA 5948
OY 88 TTGAGATTGATTTGCGTGTCTCCAGATCCATCCATTAATCCAT 130
Db 5949 TTTCGTGTGATTTGAATCTGCGCTTGCAATTTTATTACTT 5991

RESULT 9
US-09-662-250A-15/C
; Sequence 15, Application US/09662250A
; Patent No. 6368856
; GENERAL INFORMATION:
; APPLICANT: Brecht P. Monia
; APPLICANT: Jacqueline Wyatt
; TITLE OF INVENTION: ANTISENSE MODULATION OF PHOSPHORYLASE KINASE BETA EXPRESS
; FILE REFERENCE: R1S-0129
; CURRENT APPLICATION NUMBER: US/09/662,250A
; EARLIER FILING DATE: 2000-09-14
; NUMBER OF SEQ ID NOS: 102
; SEQ ID NO 15
; LENGTH: 811
; TYPE: DNA
```

ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (45)...(100)
NAME/KEY: CDS
LOCATION: (609)...(702)
US-09-662-250A-15

Query Match 6.5%; Score 32.4; DB 4; Length 811;
Best Local Similarity 58.2%; Pred. No. 2.7;
Matches 57; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 1 TTCTATGAACATAAACCACCAATATTTTCAAGCTAGTCATAATATTTACTTGAT 60
DB 356 TTCTCAAACTAACACCGACACGACCTGCTGCTTATATATACCTCAAC 297
QY 61 TCATTCGATTTATTTGTCTTACATTAATTAGATTGAT 98
DB 296 TAAACAAATATATTTTAAACACGAAAGAAATATCT 259

RESULT 10

US-08-330-108-2/c
Sequence 2, Application US/08330108
Patent No. 5795752
GENERAL INFORMATION:
APPLICANT: Smith, Kendall A.
TITLE OF INVENTION: IL-2-Stimulated Gene
TITLE OF INVENTION: Expression
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Iahive & Cockfield
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/330,108
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/104,736
FILING DATE:
APPLICATION NUMBER: US/07/796,066
FILING DATE:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 188 base pairs
TYPE: nucleic acid
STRANDEDNESS: single stranded
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: no
ANTI-SENSE: no
ORIGINAL SOURCE:
ORGANISM: human
CELL TYPE: T-cell blast
IMMEDIATE SOURCE:
LIBRARY:
CLONE: 1F5
US-08-330-108-2

Query Match 6.4%; Score 32.2; DB 1; Length 188;
Best Local Similarity 51.8%; Pred. No. 1.7;
Matches 73; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
QY 55 TTGAATTCATTCGATTTATTTGTGTACATTAATAGATTGATTCGTTCTCCAGT 114

DB 185 TTTTCTTTCTCTGATTCGTTGTTCTTATTTAGTAGATTGCAATACCAAAAG 126
QY 115 CCATCCAAATATCCATTCATTTGGCAGCCAAATCCCTTTTGTGATCATACATGAG 174
DB 125 TCCCTCGATTAGCTTTTGAATCAACCATTTGCTCATTTAGTAGTCATATACATGAT 66
QY 175 TATCTCTCACAGAACACAG 195
DB 65 TATTCATGAAACATCATG 45

RESULT 11

PCT-US92-10087-2/c
Sequence 2, Application PC/TUS9210087
GENERAL INFORMATION:
APPLICANT: Smith, Kendall A.
TITLE OF INVENTION: IL-2-Stimulated Gene
TITLE OF INVENTION: Expression
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Iahive & Cockfield
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/10087
FILING DATE: 19921118
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/796,066
FILING DATE: 20-NOVEMBER-1991
ATTORNEY/AGENT INFORMATION:
NAME: DeConti, Giulio A. Jr.
REGISTRATION NUMBER: 31,503
REFERENCE/DOCKET NUMBER: DCI-028PC
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-227-7400
TELEFAX: 617-227-5941
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 188 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single stranded
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: no
ANTI-SENSE: no
ORIGINAL SOURCE:
ORGANISM: human
CELL TYPE: T-cell blast
IMMEDIATE SOURCE:
LIBRARY:
CLONE: 1F5
PCT-US92-10087-2

Query Match 6.4%; Score 32.2; DB 5; Length 188;
Best Local Similarity 51.8%; Pred. No. 1.7;
Matches 73; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 55 TTGAATTCATTCGATTTATTTGTGTACATTAATAGATTGATTCGTTCTCCAGT 114
DB 185 TTTTCTTTCTCTGATTCGTTGTTCTTATTTAGTAGATTGCAATACCAAAAG 126
QY 115 CCATCCAAATATCCATTCATTTGGCAGCCAAATCCCTTTTGTGATCATACATGAG 174

GenCore version 5.1.6
Copyright (c) 1993-2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 15:04:06 ; Search time 166.778 seconds
(without alignments)
6765.003 Million cell updates/sec

Perfect score: 501
Sequence: 1 ttctatgaactaaagcacc.....gcaccctgttgcctttt 501

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: N_Geneseq_10100280

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	38.8	7.7	593	ABV52720
2	37.4	7.5	3212	ABL06380
3	37	7.4	11036	AA545411
4	37	7.4	11036	ABK28264
5	36.4	7.3	7441	ABK40057
6	36.4	7.3	7479	AA563344
7	36.4	7.3	2155561	ABN71527
8	36.2	7.2	924	AAT83941
9	36.2	7.2	924	AAV53383

C 10	36.2	7.2	993	22	AA536807
C 11	36.2	7.2	993	22	AA536808
C 12	36	7.2	26636	23	ABL16650
C 13	35.8	7.1	3030	24	AAK99724
C 14	35.4	7.1	3030	22	AA535662
C 15	35.4	7.1	3030	22	AA534882
C 16	35.4	7.1	3734	22	AAH54739
C 17	35.2	7.0	5105	13	AAQ20536
C 18	35.2	7.0	11172	24	ABL34053
C 19	35.2	7.0	14041	22	AAH48024
C 20	35.2	7.0	50000	21	AA646140
C 21	35	7.0	1334	21	AA593330
C 22	35	7.0	1797	21	AA593329
C 23	35	7.0	3823	21	AA593329
C 24	35	7.0	6270	24	ABL33872
C 25	35	7.0	6658	18	AAT79882
C 26	35	7.0	6700	24	ABN80305
C 27	34.8	6.9	6559	23	ABL04876
C 28	34.8	6.9	7059	15	AAQ44750
C 29	34.8	6.9	9859	24	ABL34133
C 30	34.6	6.9	220895	24	ABK84798
C 31	34.4	6.9	374	23	ABV52305
C 32	34.4	6.9	542	23	ABV52528
C 33	34.4	6.9	1365	22	AA588109
C 34	34.4	6.9	11963	23	ABL05864
C 35	34.2	6.8	6447	24	ABO66973
C 36	34.2	6.8	6447	24	ABL70547
C 37	34.2	6.8	6447	24	ABL33086
C 38	34.2	6.8	6447	24	ABL34532
C 39	34	6.8	5204	24	ABL32900
C 40	34	6.8	7903	22	AA545402
C 41	34	6.8	7903	22	ABL33362
C 42	34	6.8	7903	24	ABK28247
C 43	34	6.8	12177	24	ABL32651
C 44	34	6.8	611590	21	AAV22303
C 45	34	6.8	1664976	19	AAV21209

ALIGNMENTS

RESULT 1	ABV52720	standard; cDNA; 593 BP.
ID	ABV52720	
AC	ABV52720;	
XX		
DT	17-SEP-2002 (first entry)	
DE	Human prostate expression marker CDNA 52711.	
XX		
KW	Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;	
KW	pharmacogenomic marker; gene; ss.	
OS	Homo sapiens.	
XX		
PN	WO200106860-A2.	
XX		
PD	23-AUG-2001.	
XX		
PF	20-FEB-2001; 2001WO-US05171.	
XX		
PR	17-FEB-2000; 2000US-183119P.	
PR	16-MAR-2000; 2000US-189862P.	
PR	25-MAY-2000; 2000US-207454P.	
PR	09-JUN-2000; 2000US-211314P.	
PR	18-JUL-2000; 2000US-219007P.	
PR	13-DEC-2000; 2000US-255281P.	
XX		
PA	(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.	
XX	Schlegel R, Endege WO, Monahan JE;	
PI		

Human cardiovascular
Human cardiovascular
Drosophila melanog
DNA of the Bpm1 en
Human cardiovascular
S. epidermidis gen
S. epidermidis gen
Glutamine 2-oxo-gl
Human immune syste
Internal control B
Nucleotide sequenc
Nucleotide sequenc
Nucleotide sequenc
Nucleotide sequenc
Human immune syste
Tomato immunity 2
Human chemically m
Drosophila melanog
Soybean lipoxigena
Human immune syste
Human immune syste
Human cDNA differe
Human prostate exp
Human FLEXHT-40 nu
Drosophila melanog
Human angiogenesis
Chemically treated
Human immune syste
Human metastasis a
Human immune syste
Chemically pretrea
Human immune syste
DNA transcription
Human immune syste
Arabidopsis thalia
Methanococcus jann

PT arteriosclerosis comprising fragments of chemically modified genes
PT associated with cell cycle -
XX
PS Claim 1; SEQ ID No 116; 28pp; English.
XX
CC Sequences AAS45296-AAS45570 represent chemically pretreated genomic DNA
CC molecules associated with the cell cycle and specific PCR primers of the
CC invention. The sequences are useful for detecting the methylation state
CC of all Cpg dinucleotides in a sequence and therefore for analysing
CC associated diseases. By analysing cytosine methylations in the pretreated
CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
CC of existing diseases or the predisposition to specific diseases can be
CC ascertained. The parameters may be compared to another set of genetic
CC and/or epigenetic parameters, the differences serving as basis for
CC diagnosis and/or prognosis events which are disadvantageous to patients.
CC The sequences of the invention are useful for the diagnosis and therapy
CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,
CC aging, glomerular disease, Lewy body disease, arthritis,
CC arteriosclerosis, solid tumours and cancers.
XX
SQ Sequence 11036 BP; 3112 A; 312 C; 2270 G; 5335 T; 7 other;
Query Match 7.4%; Score 37; DB 22; Length 11036;
Best Local Similarity 60.4%; Pred. No. 5.8;
Matches 61; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
QY 23 AATATTTTCAAGCTAGTCATAAATTTACTTGATTCGATTCGATTATGCTTAC 82
DB 4240 AATTTTCAGAAAGTTTGTGATTTATTTATTAATTAAGTAATCGATTATTTTATTTA 4299
QY 83 ATAAATTAGATTGATTTCGTTCTCCCAAGTCCATCCAAAT 123
DB 4300 ATTAATAATTTTCGATTTTATTTATTTAAGTGTTTAAAT 4340
RESULT 4
ABK28264
ID ABK28264 standard; DNA; 11036 BP.
XX
AC ABK28264;
XX
DT 23-APR-2002 (first entry)
XX
DE DNA transcription associated complementary genomic DNA #69.
XX
KW DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
KW PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
KW single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
KW viral infection; Sezary syndrome; haematological disorder; tuberculosis;
KW immunological disorder; Werner syndrome; developmental disorder;
KW psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
KW neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;
KW myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
KW angiodysgenesis; congenital heart disease; HDK syndrome; gene therapy;
KW polyglutamine disorder; solid tumour.
XX
OS Unidentified.
XX
PN WO200192565-A2.
XX
PD 06-DEC-2001.
XX
PF 06-APR-2001; 2001WO-EP03973.
XX
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX

DR WPI; 2002-090046/12.
XX
XX New nucleic acids or oligomers, useful for diagnosing or treating
PT diseases associated with DNA transcription, e.g. immunological
PT disorders, Werner syndrome, psoriasis, myocardial infarction, solid
PT tumours or cancer
XX
PS Claim 1; SEQ ID No 138; 32pp; English.
XX
CC The invention relates to a nucleic acid, which comprises a segment of the
CC chemically pretreated DNA of genes associated with DNA transcription from
CC one of 346 sequences, and an oligomer, in particular an oligonucleotide
CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
CC to the chemically pretreated DNA of genes associated with DNA
CC transcription. The set of oligomer probes are useful for detecting the
CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
CC in a chemically pretreated genomic DNA. The nucleic acids are useful for
CC diagnosing or treating diseases associated with DNA transcription
CC (particularly with the methylation status), e.g. adenosine deaminase
CC deficiency, viral infection, retroviral infection, Sezary syndrome,
CC haematological disorders, immunological disorders, Werner syndrome,
CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
CC neurological disorders, neurodegenerative disorders, Waardenburg
CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
CC infarction, hypertension, angiodysgenesis, erythropoiesis, congenital heart
CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
CC or cancer. Sequences ABK28127-ABK28472 represent DNA transcription
CC associated genomic DNA molecules of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC Specification but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 11036 BP; 3112 A; 312 C; 2270 G; 5335 T; 7 other;
Query Match 7.4%; Score 37; DB 24; Length 11036;
Best Local Similarity 60.4%; Pred. No. 5.8;
Matches 61; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
QY 23 AATATTTTCAAGCTAGTCATAAATTTACTTGATTCGATTCGATTATGCTTAC 82
DB 4240 AATTTTCAGAAAGTTTGTGATTTATTTATTAATTAAGTAATCGATTATTTTATTTA 4299
QY 83 ATAAATTAGATTGATTTCGTTCTCCCAAGTCCATCCAAAT 123
DB 4300 ATTAATAATTTTCGATTTTATTTATTTAAGTGTTTAAAT 4340
RESULT 5
ABK40057/C
ID ABK40057 standard; DNA; 7441 BP.
XX
AC ABK40057;
XX
DT 21-MAY-2002 (first entry)
XX
DE Human chemically pretreated gene sequence #70 strand1.
XX
KW Human; ds; bisulphite treatment; Cpg; DNA methylation; cancer; tumour;
KW cytosinatic; ALDH6; CYP11A; CYP11B1; CYP3A3; DYPD; EPHX2; OCLN; TXNRD1;
KW UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.
XX
OS Homo sapiens.
XX
PN WO200202806-A2.
XX
PD 10-JAN-2002.
XX
PF 29-JUN-2001; 2001WO-EP07470.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX

XX Olek A, Piepenbrock C, Berlin K;
 XX WPI; 2002-154757/20.
 DR
 XX
 PT New nucleic acid, oligonucleotides and peptide nucleic acid-oligomers,
 PT useful for detecting cytosine methylation state of genes associated
 PT with pharmacogenomics and for therapy of diseases e.g. cancer -
 PS
 XX Claim 1; SEQ ID No 139; 24pp; English.

CC The invention relates to a nucleic acid comprising a sequence at
 CC least 18 bases in length of a segment of the chemically pretreated DNA
 CC of genes associated with pharmacogenomics according to one of the
 CC sequences of the genes ALDH6 (NM_000693), CYP1A1 (NM_000781), CYP1B1
 CC (NM_000497), CYP3A3 (NM_000776 and NM_017460), DPYD (NM_000110), EPHX2
 CC (NM_001979), OCLN (NM_002538), TXNRD1 (NM_003330), UGT8 (NM_003360),
 CC MRP (NM_004996, NM_019900, NM_019901, NM_019902, NM_019862, NM_019898,
 CC NM_019899) and their complementary sequences, or a sequence (S1) chosen
 CC from 87 sequences and their complements. The chemical pretreatment
 CC is bisulphite treatment to convert cytosines (but not methyl-cytosines)
 CC into uracils. Also included are an oligomer (II) in particular an
 CC oligonucleotide or a peptide nucleic acid (PNA)-oligomer, comprising in
 CC each case at least one base sequence having a length of 9 nucleotides
 CC which hybridises to or is identical to a chemically pretreated DNA of
 CC genes associated with pharmacogenomics and their complements, arranged in
 CC an array for analysing diseases associated with the methylation state
 CC (CpG) and/or detecting SNPs (single nucleotide polymorphisms)
 CC of the 87 sequences. The oligomers may also be used as PCR primers.
 CC The set of 87 nucleic acids and their complements is useful for diagnosis
 CC and therapy of solid tumours and cancer. The present sequence
 CC represents one the 87 DNA sequences or its complement.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX
 SQ Sequence 7441 BP; 2098 A; 66 C; 1948 G; 3329 T; 0 other;

Query Match 7.3%; Score 36.4; DB 24; Length 7441;
 Best Local Similarity 55.6%; Pred. No. 7.4;
 Matches 70; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

OY 9 ACTAAAGCACCAATATTTTCAAGCTAAGTCAATATATTTACTGTAATTCATTCG 68
 DB 4550 AATATATCAATTCATATTTCTTAATATTCACAAAAATTTTACTTCACTACATTAAT 4491
 OY 69 CATTATGTTGTTACATAAATAGATTGATTCGTCTCCAAAGCCATCCATATATCC 128
 DB 4490 CAATATTATCTCCAAAAAATCTATACAAATACCTTAACCCCACTTAATAATTTTAC 4431
 OY 129 ATCCAT 134
 DB 4430 ATCCAT 4425

RESULT 6
 AAS63344/C
 ID AAS63344 standard; DNA; 7479 BP.
 XX
 AC AAS63344;
 XX
 DT 29-JAN-2002 (first entry)
 XX
 XX Chemically pretreated metabolism associated gene #39.
 DE
 XX Human; cytostatic; anti-tumour; metabolism; metabolic disease; liver;
 KW Solid tumour; cancer; cytosine methylation; epigenetic; eye; kidney;
 KW single nucleotide polymorphism detection; SNP; stool; urine; lung;
 KW cerebral-spinal fluid; intestine; brain; heart; prostate; breast;
 KM DUSP2; EPHX2; QDPR; SGGH; SHMT2; SLC7A2; SLC7A4; TYMS; ds.
 XX
 *OS Homo sapiens.

XX
 PN WO200176451-A2.
 XX
 PD 18-OCT-2001.
 XX
 PF 06-APR-2001; 2001MO-EP04016.
 XX
 PR 06-APR-2000; 2000DE-1019058.
 PR 07-APR-2000; 2000DE-1019173.
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIC-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX WPI; 2002-010834/01.
 DR
 XX
 PT New nucleic acid, useful for diagnosis and therapy of metabolic
 PT disease, solid tumour and cancers, comprises segment of chemically
 PT modified genomic sequences of genes associated with metabolism -
 PS
 XX Claim 1; Page 107-109; 143pp; English.

CC The invention relates to a nucleic acid (I) comprising a sequence at
 CC least 18 bases of a segment of the chemically pretreated DNA of genes
 CC associated with metabolism such as DUSP2 (NM_004418), EPHX2 (NM_001979),
 CC ODPB (NM_000320), SGGH (NM_000199), SHMT2 (NM_005412), SLC7A2
 CC (NM_003046), SLC7A4 (NM_004173) and TYMS (NM_001071) (all
 CC undefined). (I) are useful for diagnosis and therapy of metabolic
 CC disease, solid tumours and cancers; as primer oligonucleotides for the
 CC amplification of DNA sequences, for detecting the cytosine methylation
 CC state and/or single nucleotide polymorphisms (SNPs) in a chemically
 CC treated DNA of genes associated with metabolism. An array of (I) is
 CC useful for ascertaining genetic and/or epigenetic parameters for the
 CC diagnosis and/or therapy of existing diseases or the predisposition to
 CC specific diseases by analysing cytosine methylations. The method involves
 CC chemically treating genomic DNA sample by a solution of bisulphite,
 CC hydrogen sulphite or disulphite such that cytosine bases which are
 CC unmethylated at the 5th-position are converted to uracil or another base
 CC which is dissimilar to cytosine in terms of hybridisation behaviour and
 CC amplifying fragments of the chemically pretreated genomic DNA. The
 CC genomic DNA is from cells or cellular components which contain DNA,
 CC sources of DNA comprising, for e.g. cell lines, biopsies, blood, sputum,
 CC stool, urine, cerebral-spinal fluid, tissue embedded in paraffin such as
 CC tissue from eye, intestine, kidney, brain, heart, prostate, lung, breast
 CC or liver, histologic object slides and their combinations. Genetic
 CC parameters are mutations, in particular insertions, deletions, point
 CC mutations, inversions and polymorphisms of genes associated with
 CC metabolism and sequences further required for their regulation.
 CC Epigenetic parameters are in particular cytosine methylations and
 CC further chemical modifications of DNA bases of genes associated with
 CC metabolism. Further epigenetic parameters include for e.g. the
 CC acetylation of histones which correlates with DNA methylation.
 CC AAS63306-AAS63373 represent chemically pretreated metabolism associated
 CC genes, and related primers of the invention.
 XX

SQ Sequence 7479 BP; 2111 A; 68 C; 1962 G; 3338 T; 0 other;

Query Match 7.3%; Score 36.4; DB 24; Length 7479;
 Best Local Similarity 55.6%; Pred. No. 7.4;
 Matches 70; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

OY 9 ACTAAAGCACCAATATTTTCAAGCTAAGTCAATATATTTACTGTAATTCATTCG 68
 DB 4588 AATATATCAATTCATATTTCTTAATATTCACAAATTTTACTTCACTACATTAAT 4529
 OY 69 CATTATGTTGTTACATAAATAGATTGATTCGTCTCCAAAGCCATCCATATATCC 128
 DB 4528 CAATATTATCTCCAAAAAATCTATACAAATACCTTAACCCCACTTAATAATTTTAC 4469
 OY 129 ATCCAT 134

Qy	143	CAATCCTTATTGGTCATACAT	167
Db	557	TACATCTGATTACATATACAACAT	533

RESULT	11
AAS36808/c	
ID	AAS36808 standard; DNA; 993 BP.

Human cardiovascular system antigen genomic DNA SEQ ID No 2308.

KM Cardiovascular system antigen; human; mouse; rabbit; goat; horse; cat;
 KM chicken; sheep; immunosuppressive; antiarthritic; vasotropic; dog;
 KM chemopreventive; antiproliferative; cytosstatic; cardiant; neuroprotective;
 KM cerebroprotective; neurotropic; antibacterial; virucide; fungicide; cancer;
 KM ophthalmological; vulnerary; gene therapy; autoimmune disease; neoplasm;
 KM hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;
 KM cerebrovascular disorder; nervous system disorder; bacterial infection;
 KM fungal infection; viral infection; cancer; endocrine disorder;
 KM gastrointestinal disorder; renal disorder; respiratory disorder;
 KM wound healing; skin aging; organ transplantation; tissue regeneration;
 KM anti-infectivity.

OS	Homo sapiens.
XX	~
PN	W0200155321-A2.

PD 02-AUG-2001.

PF 17-JAN-2001; 2001WO-US01340.

PR	07-JUN-2000	2000S-0209467
PR	28-JUN-2000	2000S-0214686
PR	30-JUN-2000	2000S-0215135
PR	07-JUL-2000	2000S-0216640
PR	07-JUL-2000	2000S-0216887
PR	11-JUL-2000	2000S-0217487
PR	11-JUL-2000	2000S-0217496
PR	14-JUL-2000	2000S-0218290
PR	26-JUL-2000	2000S-0218290
PR	26-JUL-2000	2000S-0218293
PR	26-JUL-2000	2000S-0220964
PR	14-AUG-2000	2000S-0224518
PR	14-AUG-2000	2000S-0224519
PR	14-AUG-2000	2000S-0225113
PR	14-AUG-2000	2000S-0225113
PR	14-AUG-2000	2000S-0225366
PR	14-AUG-2000	2000S-0225367
PR	14-AUG-2000	2000S-0225568
PR	14-AUG-2000	2000S-0225570
PR	14-AUG-2000	2000S-0225447
PR	14-AUG-2000	2000S-0225572
PR	14-AUG-2000	2000S-0225578
PR	14-AUG-2000	2000S-0225759
PR	18-AUG-2000	2000S-0225759
PR	22-AUG-2000	2000S-0226671
PR	22-AUG-2000	2000S-0226681
PR	22-AUG-2000	2000S-0226682
PR	23-AUG-2000	2000S-0227103
PR	30-SEP-2000	2000S-0228292
PR	01-SEP-2000	2000S-0229287

PR	01-SEP-2000	2000US-0229343
PR	01-SEP-2000	2000US-0229344
PR	01-SEP-2000	2000US-0229345
PR	05-SEP-2000	2000US-0229509
PR	05-SEP-2000	2000US-0229513
PR	06-SEP-2000	2000US-0230437
PR	06-SEP-2000	2000US-0230438
PR	08-SEP-2000	2000US-0231242
PR	08-SEP-2000	2000US-0231243
PR	08-SEP-2000	2000US-0231244
PR	08-SEP-2000	2000US-0231413
PR	08-SEP-2000	2000US-0232080
PR	08-SEP-2000	2000US-0232081
PR	12-SEP-2000	2000US-0232086
PR	14-SEP-2000	2000US-0232397
PR	14-SEP-2000	2000US-0232398
PR	14-SEP-2000	2000US-0232399
PR	14-SEP-2000	2000US-0232400
PR	14-SEP-2000	2000US-0232401
PR	14-SEP-2000	2000US-0233063
PR	14-SEP-2000	2000US-0233065
PR	14-SEP-2000	2000US-0233065
PR	21-SEP-2000	2000US-0234423
PR	21-SEP-2000	2000US-0234424
PR	25-SEP-2000	2000US-0234997
PR	25-SEP-2000	2000US-0234997
PR	25-SEP-2000	2000US-0234998
PR	26-SEP-2000	2000US-0235484
PR	27-SEP-2000	2000US-0235484
PR	27-SEP-2000	2000US-0235484
PR	29-SEP-2000	2000US-0236359
PR	29-SEP-2000	2000US-0236359
PR	02-OCT-2000	2000US-0236802
PR	02-OCT-2000	2000US-0237037
PR	02-OCT-2000	2000US-0237038
PR	02-OCT-2000	2000US-0237039
PR	02-OCT-2000	2000US-0237040
PR	02-OCT-2000	2000US-0239495
PR	13-OCT-2000	2000US-0239937
PR	20-OCT-2000	2000US-0240960
PR	20-OCT-2000	2000US-0241221
PR	20-OCT-2000	2000US-0241185
PR	20-OCT-2000	2000US-0241186
PR	20-OCT-2000	2000US-0241187
PR	20-OCT-2000	2000US-0241808
PR	20-OCT-2000	2000US-0241809
PR	01-NOV-2000	2000US-0241826
PR	01-NOV-2000	2000US-0244617
PR	08-NOV-2000	2000US-0246474
PR	08-NOV-2000	2000US-0246475
PR	08-NOV-2000	2000US-0246476
PR	08-NOV-2000	2000US-0246477
PR	08-NOV-2000	2000US-0246478
PR	08-NOV-2000	2000US-0246523
PR	08-NOV-2000	2000US-0246524
PR	08-NOV-2000	2000US-0246524
PR	08-NOV-2000	2000US-0246525
PR	08-NOV-2000	2000US-0246527
PR	08-NOV-2000	2000US-0246527
PR	08-NOV-2000	2000US-0246528
PR	08-NOV-2000	2000US-0246532
PR	08-NOV-2000	2000US-0246609
PR	08-NOV-2000	2000US-0246610
PR	08-NOV-2000	2000US-0246611
PR	08-NOV-2000	2000US-0246613
PR	17-NOV-2000	2000US-0249207
PR	17-NOV-2000	2000US-0249208
PR	17-NOV-2000	2000US-0249209
PR	17-NOV-2000	2000US-0249210
PR	17-NOV-2000	2000US-0249211
PR	17-NOV-2000	2000US-0249212

RESULT 13
AAK9724/C
ID AAK9724 standard; DNA: 3030 BP.
XX
AC AAK9724;
XX
DT 11-JUL-2002 (first entry)
XX
DE DNA of the BpmI endonuclease gene (BpmIRM).
XX
KW BpmI restriction endonuclease; Bacillus pumilus; recombinant expression;
KM cleaving DNA; cloning; gene characterisation; methylase; BpmIRM;
KM BpmI endonuclease gene; gene; ds.
XX
OS Bacillus pumilus.
XX
FH Key location/Qualifiers
FT CDS 1..3030
FT /tag- a
FT /product- "Protein of the BpmI endonuclease gene
FT (BpmIRM)"
XX
PN EPI199365-A2.
XX
PD 24-APR-2002.
XX
PE 10-OCT-2001; 2001EP-0203826.
XX
PR 20-OCT-2000; 2000US-0693146.
XX
PA (NEW) NEW ENGLAND BIO LABS INC.
XX
PI Xu S, Xiao J, Zhu Z;
XX
DR WPI: 2002-354206/39.
DR P-PSDB: AAO20949.
XX
PT New isolated DNA encoding BpmI restriction endonuclease, useful for
PT cleaving DNA and for the associated methylase
XX
PS Disclosure; Fig 3; 33pp; English.
XX
CC The invention relates to an isolated DNA segment encoding the BpmI
CC restriction endonuclease and isolated from Bacillus pumilus, New England
CC Biolabs Collection No. 711. The isolated DNA is used for recombinant
CC expression of BpmI, and is useful as a reagent for cleaving DNA, and for
CC cloning or gene characterisation. The associated methylase, BpmIM, can
CC be used similarly. This polynucleotide sequence represents the DNA of the
CC BpmI endonuclease gene (BpmIRM) of the invention.
XX
SQ Sequence 3030 BP; 1137 A; 405 C; 572 G; 916 T; 0 other;
Query Match 7.1%; Score 35.8; DB 24; Length 3030;
Best Local Similarity 52.3%; Pred. No. 8;
Matches 73; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
QY 4 TATGACTAAAGACACATATTTTCAGAGTCAATATTTACTGCAATCA 63
DB 1339 TAATATCAACGCAACAATAATTTCTCCAGATATTTCTTTAGTCAAA 1280
QY 64 ATTGCATTTATGTTACATTAATTTAGATTGATTTGGTTCTCCAGTCCATCCAT 123
DB 1279 TTACCCCTTCTTCAGTTTCAATATTTGATCATGATCTCTCTTATTAAT 1220
QY 124 TATGATCATTTGGGACGCCCAATCCCTTA 154
DB 1219 TTTCATTAATCTTCTCTACTAGAAAGTCA 1189

XX
AC AAK35662;
XX
DT 17-DEC-2001 (first entry)
XX
DE Human cardiovascular system antigen cdNA polynucleotide SEQ ID No 547.
XX
KW Cardiovascular system antigen; human; mouse; rabbit; goat; horse; cat;
KW chicken; sheep; immunosuppressive; antiarthritic; vasotropic; dog;
KW antineumatic; antiproliferative; cytostatic; cardiant; neuroprotective;
KW cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;
KW ophthalmological; vulnery; gene therapy; autoimmune disease; neoplasm;
KW hyperproliferative disorder; breast; liver; cardiovascular disorder; ss;
KW cerebrovascular disorder; nervous system disorder; bacterial infection;
KW fungal infection; viral infection; ocular disorder; endocrine disorder;
KW gastrointestinal disorder; renal disorder; respiratory disorder;
KW wound healing; skin aging; organ transplantation; tissue regeneration;
KW anti-infertility.
XX
OS Homo sapiens.
XX
PN WO200155321-A2.
XX
PD 02-AUG-2001.
XX
PE 17-JAN-2001; 2001WO-US01340.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 14-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227189.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 03-SEP-2000; 2000US-0229509.
PR 03-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.

Job time : 169.778 secs

ID AAH54882 standard; DNA: 3030 BP.
XX
AC .AAH54882;
XX
DT 03-SEP-2001 (first entry)
XX
DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:4246.
XX
KM Staphylococcus epidermidis SRI strain; infection; diagnosis;
XX vaccination; endocarditis; ds.
XX
OS Staphylococcus epidermidis.
XX
PM W0200134809-A2.
XX
PD 17-MAY-2001.
XX
PF 09-NOV-2000; 2000WO-US30782.
XX
PR 09-NOV-1999; 99US-0164258.
XX
PA (GLAX) GLAXO GROUP LTD.
XX
PI Kimmerly WJ;
XX
DR WPI: 2001-316495/33.
XX
PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
XX useful for vaccinating against infections, e.g. endocarditis -
XX
PS Claim 8; Page 1968-1969; 2188pp; English.
XX
CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX
SO Sequence 3030 BP: 1135 A; 419 C; 558 G; 918 T; 0 other;

Query Match 7.1%; Score 35.4; DB 22; Length 3030;
Best Local Similarity 53.2%; Pred. No. 10;
Matches 75; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

OY 15 AGCACCAATATTTCAGAGCTAGTCATTAATATTACTTGATGATCAATTCGATTTAT 74
DB 1441 AACTGCACAAATTTTATTAAGATGATCTCATCTAGTACTAGATTGCGCAAAATPAG 1500
75 TGTGTACATTAATTAATTTGATTTGCTGTCACAGTCATCATTCATTCATCCAT 134
OY 1501 TAGGTAAATTAATAACGAGGTGACTTGATGATCATCATGACTATCTAATTTACTGTAAT 1560
DB 135 TTGGCAGCCAAATCCTCTTAT 155
1561 TTCTTATCTAATAAGTTAAT 1581

Search completed: July 4, 2003, 15:25:04

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 15:08:57 ; Search time 1508.67 Seconds

(without alignments)
9664.474 Million cell updates/sec

US-10-083-853b-2_COPY_18900_19400

Sequence: 1 ttctatgaactaaagcacc.....gcatccctgttcttttt 501

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

GenEmbl: *
1: gb_ba: *
2: gb_htg: *
3: gb_in: *
4: gb_ov: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_or: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vl: *
30: em_htg_hum: *
31: em_htg_in: *
32: em_htg_other: *
33: em_htg_mus: *
34: em_htg_pln: *
35: em_htg_rtd: *
36: em_htg_man: *
37: em_htg_vfl: *
38: em_sy: *
39: em_htg_hum: *
40: em_htg_mus: *
41: em_htg_other: *

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	501	100.0	180648	9 AP003043	AP003043 Homo sapi
2	501	100.0	182403	2 AC024052	AC024052 Homo sapi
3	65	13.0	160345	2 AC067870	AC067870 Homo sapi
4	43	8.6	176784	2 AC016632	AC016632 Homo sapi
5	43	8.6	182126	2 AC016639	AC016639 Homo sapi
6	41.4	8.3	193087	2 AC010779	AC010779 Mus muscu
7	40.4	8.1	149628	2 AC080186	AC080186 Homo sapi
8	40.4	8.1	172868	9 AC019140	AC019140 Homo sapi
9	40.4	8.1	176951	2 AL845340	AL845340 Mus muscu
10	40.2	8.0	74673	2 AC005683	AC005683 Homo sapi
11	40.2	8.0	227054	9 AF172277	AF172277 Homo sapi
12	40	8.0	55470	2 AC116989	AC116989 Dictyoste
13	40	8.0	150950	2 AC119464	AC119464 Rattus no
14	39.8	7.9	103138	2 AC112117	AC112117 Rattus no
15	39.8	7.9	139999	8 AC018727	AC018727 Oryza sat
16	39.8	7.9	163166	9 AL356632	AL356632 Human DNA
17	39.8	7.9	171014	10 AL671915	AL671915 Mouse DNA
18	39.6	7.9	188993	2 AC109656	AC109656 Rattus no
19	39.6	7.9	3457	3 AF310886	AF310886 Dictyoste
20	39.6	7.9	4197	3 AF163835	AF163835 Dictyoste
21	39.2	7.8	107341	10 AL805932	AL805932 Mouse DNA
22	39.2	7.8	138397	2 AC095231	AC095231 Rattus no
23	39.2	7.8	148391	2 AC034174	AC034174 Homo sapi
24	39.2	7.8	159012	2 AC023145	AC023145 Homo sapi
25	39.2	7.8	159972	2 AC107810	AC107810 Mus muscu
26	39.2	7.8	160250	9 AC087173	AC087173 Homo sapi
27	39.2	7.8	160454	9 AC087366	AC087366 Homo sapi
28	39.2	7.8	163717	2 AC127281	AC127281 Mus muscu
29	39.2	7.8	192404	9 CENS1DWO	AL136296 Human chr
30	39.2	7.8	194985	10 AC002406	AC002406 Mouse chr
31	39	7.8	61513	9 AL353610	AL353610 Human DNA
32	39	7.8	160324	9 AC079117	AC079117 Homo sapi
33	38.8	7.7	43834	3 AF045643	AF045643 Caenorhab
34	38.8	7.7	66679	2 AC118587	AC118587 Mus muscu
35	38.8	7.7	161495	2 AC120893	AC120893 Rattus no
36	38.8	7.7	171668	2 AC112975	AC112975 Mus muscu
37	38.8	7.7	187023	2 AC006882	AC006882 Caenorhab
38	38.6	7.7	67247	2 AC121357	AC121357 Mus muscu
39	38.4	7.7	104333	9 AL133295	AL133295 Human DNA
40	38.4	7.7	138159	2 AP002742	AP002742 Homo sapi
41	38.4	7.7	164301	2 AC018671	AC018671 Homo sapi
42	38.4	7.7	164513	9 AC093262	AC093262 Homo sapi
43	38.4	7.7	164996	2 AC087372	AC087372 Homo sapi
44	38.4	7.7	197665	9 AC108456	AC108456 Homo sapi
45	38.2	7.6	27908	3 CETO4H1	278200 Caenorhabd1

ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	AP003043	180648 bp	DNA	linear	PRI 18-SEP-2001
AP003043	LOCUS	Homo sapiens genomic DNA, chromosome 11q clone:RP11-617B3, complete					
AP003043	DEFINITION	Homo sapiens genomic DNA, chromosome 11q clone:RP11-617B3, complete					
AP003043	VERSION	GI:15637094					
AP003043	KEYWORDS	HTG.					
AP003043	SOURCE	Homo sapiens DNA, clone:RP11-617B3.					
AP003043	ORGANISM	Homo sapiens					
AP003043	REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.					
AP003043	AUTHORS	Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P., Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.					

Pred. No. is the number of results predicted by chance to have a

TITLE Homo sapiens genomic DNA
JOURNAL Published Only in Database (2000)
REFERENCE 2 (bases 1 to 180648)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Tokoko,Y., Matsuabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (12-DEC-2000) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT On Sep 17, 2001 this sequence version replaced gi:11862945.
FEATURES
source
1.180648
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-617B3"
BASE COUNT 56234 a 35375 c 33753 g 55286 t
ORIGIN
Query Match 100.0%; Score 501; DB 9; Length 180648;
Best Local Similarity 100.0%; Pred. No. 4,5e-124; Indels 0; Gaps 0;
Matches 501; Conservative 0; Mismatches 0;
QY 1 TTCTATGAACCTAAGACACCAACATATTTTCAAGCTAAGTCAATATTTACTTGAAT 60
DB 58769 TTCTATGAACCTAAGACACCAACATATTTTCAAGCTAAGTCAATATTTACTTGAAT 58828
QY 61 TCATTCGCATTTTGTGTACATTAATTAATTAATTTGCTTCTCCAGTCCATCC 120
DB 58829 TCATTCGCATTTTGTGTACATTAATTAATTAATTTGCTTCTCCAGTCCATCC 58888
QY 121 AATTATCCATTCATTTGGGACCAAAATCTTATTTGTCATACATACATGATATCTT 180
DB 58889 AATTATCCATTCATTTGGGACCAAAATCTTATTTGTCATACATACATGATATCTT 58948
QY 181 CTCACAGAAGACAGTGAAGAGCAAGTCTTTGTCTGCACAGAACCCACCCCTTTC 240
DB 58949 CTCACAGAAGACAGTGAAGAGCAAGTCTTTGTCTGCACAGAACCCACCCCTTTC 59008
QY 241 CAATGACGCCCTGTGTGACGACGACCAACAGAAAGTCCATGATGTTAGGACG 300
DB 59009 CAATGACGCCCTGTGTGACGACGACCAACAGAAAGTCCATGATGTTAGGACG 59068
QY 301 TTATCAACATGTTCAATTAACCTAGCCCTGTGAATGTCTTAGGCTTAACGTAAC 360
DB 59069 TTATCAACATGTTCAATTAACCTAGCCCTGTGAATGTCTTAGGCTTAACGTAAC 59128
QY 361 TCGTTAAGAGAAAGAGAAAAAAGCCCTTTGGGACAGAGGTCAAAAGTCAAGACAGT 420
DB 59129 TCGTTAAGAGAGAAAGAGAAAAAAGCCCTTTGGGACAGAGGTCAAAAGTCAAGACAGT 59188
QY 421 GTCAGTAGAATGATCTAAGAGAAATTTGAGAGAAAGTGGACCATAGTCCATAC 480
DB 59189 GTCAGTAGAATGATCTAAGAGAAATTTGAGAGAAAGTGGACCATAGTCCATAC 59248
QY 481 CGCATCCCTGTTTGTCTTTT 501
DB 59249 CGCATCCCTGTTTGTCTTTT 59269
RESULT 2
AC024052 182403 bp DNA linear HTG 17-AUG-2000
LOCUS Homo sapiens chromosome 11 clone RP11-617B3, WORKING DRAFT
DEFINITION
SEQUENCE, 7 unordered pieces.
ACCESSION AC024052
VERSION AC024052.3 GI:9838295
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 182403)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 182403)
Waterston,R.H.
Direct Submission
Submitted (20-FEB-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
On Aug 17, 2000 this sequence version replaced gi:1109658.
Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
Project Information
Center project name: H_NH0617B03
Summary Statistics
Sequencing vector: M13; 100%
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 178574 bases at least Q40
Consensus quality: 179480 bases at least Q30
Consensus quality: 179940 bases at least Q20
Insert size: 198000; agarose-fp
Insert size: 183143; sum-of-contigs
Quality coverage: 5.39 in Q20 bases; agarose-fp
Quality coverage: 5.87 in Q20 bases; sum-of-contigs
NOTE: This is a 'working draft' sequence. It currently consists of 7 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
1 1666: contig of 1666 bp in length
* 1667 1766: gap of unknown length
* 1767 9133: contig of 7367 bp in length
* 9134 9233: gap of unknown length
* 9234 20663: contig of 11430 bp in length
* 20664 20763: gap of unknown length
* 20764 35927: contig of 15164 bp in length
* 35928 36027: gap of unknown length
* 36028 61610: contig of 25583 bp in length
* 61611 61710: gap of unknown length
* 61711 96167: contig of 34457 bp in length
* 96168 96267: gap of unknown length
* 96268 182403: contig of 86136 bp in length.
FEATURES
source
1.182403
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/clone="RP11-617B3"
1.1666
/note="assembly_name:Contig2"
1767.9133
/note="assembly_name:Contig3
clone_end:sp6
vector_side:left"
9234.20663
/note="assembly_name:Contig4"
20764.35927
/note="assembly_name:Contig5"
36028.61610
/note="assembly_name:Contig6"

```

misc_feature      61711..96167
                  /note="assembly_name:Contig7"
misc_feature      96268..182403
                  /note="assembly_name:Contig8
                  vector_side:left"
BASE COUNT      55923 a 35190 c 34519 g 56169 t 602 others
ORIGIN
Query Match      100.0%; Score 501; DB 2; Length 182403;
Best Local Similarity 100.0%; Pred. No. 4,3e-124;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TTCTAGACATTAAGACACACAAATTTTCAAGCTATGATTAATTTACTTGAT 60
DB 155014 TTCTATGACATTAAGACACACAAATTTTCAAGCTATGATTAATTTACTTGAT 155073
OY 61 TCAATTCGATTAATTTGTGTACATTAATTTAGATTTGTTGTTTCCAGTCCATCC 120
DB 155074 TCAATTCGATTAATTTGTGTACATTAATTTAGATTTGTTGTTTCCAGTCCATCC 155133
OY 121 AATTATCCATTTTGGACGCAAAATCCTTATTTGTGATATCATGAGATCTT 180
DB 155134 AATTATCCATTTTGGACGCAAAATCCTTATTTGTGATATCATGAGATCTT 155193
OY 181 CTCACAGAACACAGATTAGAACAGTCTTTTGTCTGACAGAACCCACCCTTTCC 240
DB 155194 CTCACAGAACACAGATTAGAACAGTCTTTTGTCTGACAGAACCCACCCTTTCC 155253
OY 241 CAATGACAGCCCTTGTGTGACACGACGACACACAGAAAGAAATCCCATGATTTAGAGCAG 300
DB 155254 CAATGACAGCCCTTGTGTGACACGACGACACACAGAAAGAAATCCCATGATTTAGAGCAG 155313
OY 301 TTATCACCATTTTATCTTAACTTACGCTGTAATTTAGCTTAAGCTTAACGTGAAC 360
DB 155314 TTATCACCATTTTATCTTAACTTACGCTGTAATTTAGCTTAAGCTTAACGTGAAC 155373
OY 361 TGTCTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
DB 155374 TGTCTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 155433
OY 421 GTCACTAGATGATCTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
DB 155434 GTCACTAGATGATCTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 155493
OY 481 CGCATCCCTGTTTGTCTTTT 501
DB 155494 CGCATCCCTGTTTGTCTTTT 155514

RESULT 3
LOCUS      AC067870      160345 bp      DNA      linear      HTG 23-JUN-2000
DEFINITION      Homo sapiens chromosome 11 clone RP11-475J2 map 11, WORKING DRAFT
ACCESSION      AC067870
VERSION      AC067870.3 GI:8671974
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 160345)
AUTHORS      Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE      Homo sapiens chromosome 11, clone RP11-475J2
JOURNAL      Unpublished
REMARKS      2 (bases 1 to 160345)
AUTHORS      Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bega,F.,
Boguslavsky,L., Boukhalter,B., Brown,A., Burkett,G.,
Camporipano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,

```

```

Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Kartas,A.,
Klein,J., Laroque,K., Lamazares,R., Landers,T., Lehotzky,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Margulis,N.,
McCarthy,M., McEwan,P., McGuck,A., McKernan,K., McPheters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenka,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisanil,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 23, 2000 this sequence version replaced gi:8099852.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-----
Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIDR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-----
Project Information
Center project name: L8810
Center clone name: 475_J-2
-----
Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 151812 bases at least Q40
Consensus quality: 156075 bases at least Q30
Consensus quality: 157717 bases at least Q20
Insert size: 163000; agarose-fp
Insert size: 158645; sum-of-coverage
Quality coverage: 4.9 in Q20 bases; agarose-fp
Quality coverage: 5.0 in Q20 bases; sum-of-coverage
-----
NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
-----
1 946: contig of 946 bp in length
* 947 1046: gap of 100 bp
* 1047 2803: contig of 1757 bp in length
* 2804 2903: gap of 100 bp
* 2904 4474: contig of 1571 bp in length
* 4475 4574: gap of 100 bp
* 4575 8112: contig of 3538 bp in length
* 8113 8212: gap of 100 bp
* 8213 11442: contig of 3320 bp in length
* 11443 11542: gap of 100 bp
* 11543 14027: contig of 2485 bp in length
* 14028 14127: gap of 100 bp
* 14128 17482: contig of 3355 bp in length
* 17483 17582: gap of 100 bp
* 17583 20007: contig of 2425 bp in length
* 20008 20107: gap of 100 bp
* 20108 26548: contig of 6441 bp in length
* 26549 26648: gap of 100 bp
* 26649 34423: contig of 7775 bp in length
* 34424 34523: gap of 100 bp
* 34524 41559: contig of 7036 bp in length
* 41560 41659: gap of 100 bp
* 41660 50886: contig of 9227 bp in length

```


QY 2 TCTATGACAAAGAACCCACCAATATTTCCTAAGCAGTGTGAATAATTTACTTGAT 61
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

D0 123123 TCTAAAATTAATTAACCAACAATTAATTAATCATATCTTAAACACTTAATTTACTTAGTAT 123064
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 62 CAATTCGCATTATTTGGTTACATAAATTAATTAATTTAGATTGATTT 100
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

D0 123063 CATTAGTAGTATCATGAGTTTACAAATCTCTGAATTTGATTT 123025

RESULT 5
AC016639 182126 bp DNA linear HTG 18-JUL-2000
AC016639/c Homo sapiens chromosome 5 clone RP11-412L4, WORKING DRAFT SEQUENCE,
LOCUS DEFINITION 7 ordered pieces.
ACCESSION AC016639
VERSION AC016639.5 GI:7711585
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 182126)
TITLE DOE Joint Genome Institute.
JOURNAL Sequencing of Human Chromosome 5
REFERENCE Unpublished
2 (bases 1 to 182126)
DOE Joint Genome Institute.
DIRECT SUBMISSION Submitted (04-DEC-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On May 6, 2000 this sequence version replaced gi:7710162.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 564841
Center clone name: RPCI-11_412L4

Summary Statistics
Consensus quality: 175214 bases at least Q40
Consensus quality: 180404 bases at least Q30
Consensus quality: 181112 bases at least Q20
Estimated insert size: 178000; pulse-field gel estimation
Estimated insert size: 181876; sum-of-contigs estimation
Quality coverage: 6.41 in Q20 bases; pulse-field gel estimation
Quality coverage: 6.28 in Q20 bases; sum-of-contigs estimation.
NOTE: This is a 'working draft' sequence. It currently
consists of 7 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces
is believed to be correct as given, however the sizes
of the gaps between them are based on estimates that have
been provided by the submitter.
This sequence will be replaced
by the finished sequence as soon as it is available and
the accession number will be preserved.
1 48575: config of 48575 bp in length
48576 48575: gap of unknown length
48676 62810: gap of 14135 bp in length
62811 62910: gap of unknown length
62911 81913: config of 19003 bp in length
81914 82013: gap of unknown length
82014 92166: config of 10153 bp in length
92167 92266: gap of unknown length
92267 167478: config of 75212 bp in length
167479 167578: gap of unknown length
167579 178445: config of 10867 bp in length
178446 178546: gap of unknown length
178546 182126: config of 3581 bp in length.
Location/Qualifiers
1. 182126 "Homo sapiens"
/organism="Homo sapiens"

Query Match 8.6%; Score 43; DB 2; Length 182126;
Best Local Similarity 64.6%; Pred. No. 0.6;
Matches 64; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 2 TCTATGACAAAGAACCCACCAATATTTCCTAAGCAGTGTGAATAATTTACTTGAT 61
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

D0 123687 TCTAAAATTAATTAACCAACAATTAATTAATCATATCTTAAACACTTAATTTACTTAGTAT 1236
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 62 CAATTCGCATTATTTGGTTACATAAATTAATTAATTTAGATTGATTT 100
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

D0 123627 CATTAGTAGTATCATGAGTTTACAAATCTCTGAATTTGATTT 123589

RESULT 6
AC101779 193087 bp DNA linear HTG 21-AUG-2002
AC101779 Mus musculus clone RP24-132P13, WORKING DRAFT SEQUENCE, 31
LOCUS DEFINITION 7 unordered pieces.
ACCESSION AC101779
VERSION AC101779.2 GI:22380791
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULFILL.
SOURCE house mouse.
ORGANISM Mus musculus.

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS 1 (bases 1 to 193087)
TITLE Birren,B., Nusbaum,C. and Lander,E.
JOURNAL Mus musculus, clone RP24-132P13
REFERENCE Unpublished
2 (bases 1 to 193087)
Birren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N.,
Anderson,S., Barina,N., Bastien,V., Boguslavsky,L., Boukhalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Chapel,Y., Colangelo,M., Collins,S., Collamore,A., Cook,A.,
Cooke,P., Deatellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferrelira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Gardya,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Hearford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamet,A., Karatas,A., Kells,C., Labrecque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Margulis,N., Matthews,C.,
McCarthy,M., McKean,P., McKernan,K., McPheters,R., Meldrim,J.,
Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Olivier,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnupack,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,L., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission
JOURNAL Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 193087)

REFERENCE Birren,B., Nusbaum,C., Lander,E., All,A., Allen,N., Anderson,S.,
Barina,N., Bastien,V., Boguslavsky,L., Boukhalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collamore,A.,
Cook,A., Cooke,P., Deatellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferrelira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamet,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mienga,V.,

TITLE
JOURNAL
COMMENT

Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Plunkhant, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Schumann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 21, 2002 this sequence version replaced g1:17060554.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: L17392
Center clone name: 132.F.13

Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 182965 bases at least Q40
Consensus quality: 186385 bases at least Q30
Consensus quality: 187886 bases at least Q20
Insert size: 182000; agarose-fp
Insert size: 190087; sum-of-contigs
Quality coverage: 6.1 in Q20 bases; agarose-fp
Quality coverage: 5.9 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 31 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
* 697: contig of 697 bp in length
* 698 797: gap of 100 bp
* 798 1442: contig of 645 bp in length
* 1443 1542: gap of 100 bp
* 1543 2213: contig of 671 bp in length
* 2214 2313: gap of 100 bp
* 2314 3123: contig of 810 bp in length
* 3124 3223: gap of 100 bp
* 3224 3698: contig of 475 bp in length
* 3699 3798: gap of 100 bp
* 3799 4425: contig of 627 bp in length
* 4426 4525: gap of 100 bp
* 4526 5160: contig of 635 bp in length
* 5161 5260: gap of 100 bp
* 5261 5949: contig of 689 bp in length
* 5950 6049: gap of 100 bp
* 6050 6708: contig of 659 bp in length
* 6709 6808: gap of 100 bp
* 6809 7498: contig of 690 bp in length
* 7499 7598: gap of 100 bp
* 7599 8497: contig of 899 bp in length
* 8498 8597: gap of 100 bp
* 8598 9482: contig of 885 bp in length
* 9483 9582: gap of 100 bp
* 9583 11049: contig of 1467 bp in length
* 11050 11149: gap of 100 bp
* 11150 12144: contig of 995 bp in length
* 12145 12244: gap of 100 bp
* 12245 14273: contig of 2029 bp in length
* 14274 14373: gap of 100 bp

14374 15864: contig of 1491 bp in length
* 15865 15964: gap of 100 bp
* 15965 17346: contig of 1382 bp in length
* 17347 17446: gap of 100 bp
* 17447 21313: contig of 3867 bp in length
* 21314 21413: gap of 100 bp
* 21414 25335: contig of 3922 bp in length
* 25336 25435: gap of 100 bp
* 25436 32017: contig of 6582 bp in length
* 32018 32117: gap of 100 bp
* 32118 44118: contig of 12001 bp in length
* 44119 44218: gap of 100 bp
* 44219 54232: contig of 10014 bp in length
* 54233 54332: gap of 100 bp
* 54333 63775: contig of 9443 bp in length
* 63776 63875: gap of 100 bp
* 63876 73299: contig of 9424 bp in length
* 73300 73399: gap of 100 bp
* 73400 85938: contig of 12539 bp in length
* 85939 86038: gap of 100 bp
* 86039 97746: contig of 11708 bp in length
* 97747 97846: gap of 100 bp
* 97847 113732: contig of 15886 bp in length
* 113733 113832: gap of 100 bp
* 113833 128811: contig of 14979 bp in length
* 128812 128911: gap of 100 bp
* 128912 149852: contig of 20941 bp in length
* 149853 149952: gap of 100 bp
* 149953 169195: contig of 19243 bp in length
* 169196 169295: gap of 100 bp
* 169296 193087: contig of 23792 bp in length.

FEATURES

source

misc_feature
1. 697
/note="assembly-fragment"
misc_feature
798. 1442
/note="assembly-fragment"
misc_feature
1543. 2213
/note="assembly-fragment"
misc_feature
2314. 3123
/note="assembly-fragment"
misc_feature
3224. 3698
/note="assembly-fragment"
misc_feature
3799. 4425
/note="assembly-fragment"
misc_feature
4526. 5160
/note="assembly-fragment"
misc_feature
5261. 5949
/note="assembly-fragment"
misc_feature
6050. 6708
/note="assembly-fragment"
misc_feature
6809. 7498
/note="assembly-fragment"
misc_feature
7599. 8497
/note="assembly-fragment"
misc_feature
8598. 9482
/note="assembly-fragment"
misc_feature
9583. 11049
/note="assembly-fragment"
misc_feature
11150. 12144
/note="assembly-fragment"
misc_feature
12245. 14273
/note="assembly-fragment"
misc_feature
13744. 15864
/note="assembly-fragment"
misc_feature
15965. 17346
/note="assembly-fragment"
misc_feature
17447. 21313
/note="assembly-fragment"


```
repeat_region      18589. .18690
/rpt_family="L2"
repeat_region      18691. .18989
/rpt_family="Alu"
repeat_region      18990. .21365
/rpt_family="L2"
repeat_region      21374. .21472
/rpt_family="MIR"
repeat_region      21732. .21812
/rpt_family="L2"
repeat_region      22068. .22615
/rpt_family="L2"
repeat_region      23058. .23403
/rpt_family="MIR"
repeat_region      23672. .23767
/rpt_family="L2"
repeat_region      25785. .26082
/rpt_family="Alu"
repeat_region      29162. .29579
/rpt_family="ERV1"
repeat_region      32263. .32394
/rpt_family="L2"
repeat_region      32464. .32611
/rpt_family="MIR"
repeat_region      32622. .32694
/rpt_family="Alu"
repeat_region      32767. .32977
/rpt_family="L2"
misc_feature       34502. .34765
/note="similar to EST A1681456 (NID:g4891638) tx37e07.x1"
repeat_region      44600. .44668
/rpt_family="MALR"
repeat_region      44878. .45052
/rpt_family="MER103"
repeat_region      45370. .45614
/rpt_family="MIR"
repeat_region      45918. .46175
/rpt_family="L2"
repeat_region      45943. .46339
/rpt_family="Alu"
misc_feature       46457. .46783
/note="similar to EST H75635 (NID:g1049957) yu07b07.s1"
repeat_region      47544. .47831
/note="similar to EST H75706 (NID:g1049649) yu07b07.r1"
repeat_region      47832. .47936
/rpt_family="Alu"
repeat_region      48162. .48306
/rpt_family="MIR"
repeat_region      49505. .49642
/rpt_family="MIR"
repeat_region      49828. .50351
/rpt_family="MIR"
repeat_region      50787. .51007
/rpt_family="ERV1"
repeat_region      52402. .52491
/rpt_family="MER1_type"
repeat_region      52499. .52563
/rpt_family="MER53"
repeat_region      53821. .53906
/rpt_family="MER53"
repeat_region      53940. .54166
/rpt_family="L2"
repeat_region      54152. .54522
/rpt_family="L2"
repeat_region      54669. .54979
/rpt_family="L2"
repeat_region      54669. .54979
/rpt_family="Alu"
repeat_region      55513. .56885
/rpt_family="MALR"
repeat_region      56886. .58488
/rpt_family="MALR"
repeat_region      58491. .58792
/rpt_family="MALR"
repeat_region      58793. .59095
/rpt_family="MALR"
```

```
repeat_region      59096. .59166
/rpt_family="Alu"
repeat_region      59096. .59166
/rpt_family="MALR"

Query Match      8.1%; Score 40.4; DB 9; Length 172868;
Best Local Similarity 54.8%; Pred. No. 3;
Matches 80; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Qy      19 CCNCAATATTTTCACAACTAGTCATTAATATTTACTTGAATTCATTCATTCATTTGTC 78
Db      78630 CCAATATTTTTCATGCTCATATTAATGAGAACATGAGTATTTGCTTCCTGTC 78689
Qy      79 TTACATTAATTTAGATTGATTGCTGCTCTCAAGTCATTCATTAATTCATTCATTTGG 138
Db      78690 TCTGACTATTTTATTTATACATATATGCTCCAGTGTCATTCATTCATTCATTCAT 78749
Qy      139 CAGCCAAATCCCTTATTTGTCATTA 164
Db      78750 GGATTTTATTTATTTTCATGCGTGAANA 78775

RESULT 9
AL845340      176951 bp      DNA      linear      HTG 09-AUG-2002
LOCUS      MUS musculus chromosome 4 clone RP23-169D3, *** SEQUENCING IN
DEFINITION      PROGRESS ***, 18 unordered pieces.
ACCESSION      AL845340
VERSION      AL845340.1 GI:22205035
KEYWORDS      HTG; HTGS-PHASE1.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1..176951)
AUTHORS      Sims, S.
TITLE      Direct Submission
JOURNAL      Submitted (05-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

COMMENT      ----- Genome Center
Center: UK Medical Research Council
Center code: UK-MRC
Web site: http://mrseq.har.mrc.ac.uk
Contact: mouse@har.mrc.ac.uk
----- Project Information
Center project name: BM169D3
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 10% of reads
Consensus quality: 164062 bases at least Q40
Consensus quality: 169446 bases at least Q30
Consensus quality: 172590 bases at least Q20
Insert size: 175251; sum-of-contrigs
Insert size: 180267; 5.0% error; agarose-1p
Quality coverage: 3.18x in Q20 bases; sum-of-contrigs Quality
coverage: 3.18x in Q20 bases; agarose-1p
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1      2625: contig of 2625 bp in length
*      2626 2725: gap of 100 bp
*      2726      5168: contig of 2443 bp in length
*      5169 5268: gap of 100 bp
*      5269      13750: contig of 8482 bp in length
*      13751 13850: gap of 100 bp
*      13851      17241: contig of 3391 bp in length
*      17242 17341: gap of 100 bp
```

```
FEATURES
source
* 17342 25111: contig of 7770 bp in length
* 25112 25211: gap of 100 bp
* 25212 33019: contig of 7808 bp in length
* 33020 33119: gap of 100 bp
* 33120 50216: contig of 17097 bp in length
* 50217 50316: gap of 100 bp
* 50317 62761: contig of 12445 bp in length
* 62762 62861: gap of 100 bp
* 62862 84579: contig of 21718 bp in length
* 84580 84679: gap of 100 bp
* 84680 98564: contig of 13885 bp in length
* 98565 98664: gap of 100 bp
* 98665 101753: contig of 3089 bp in length
* 101754 101853: gap of 100 bp
* 101854 114841: contig of 12988 bp in length
* 114842 122333: contig of 7392 bp in length
* 122334 122433: gap of 100 bp
* 122434 130729: contig of 8296 bp in length
* 130730 130829: gap of 100 bp
* 130830 135670: contig of 4841 bp in length
* 135671 135770: gap of 100 bp
* 135771 141713: contig of 5943 bp in length
* 141714 141813: gap of 100 bp
* 141814 151388: contig of 9775 bp in length
* 151389 151688: gap of 100 bp
* 151689 176951: contig of 25263 bp in length.
Location/Qualifiers
1. 176951
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="4"
/clone_lib="RPC1-23"
1. 2625
/note="assembly_fragment:01383
fragment_chain:1"
misc_feature
2726..5168
/note="assembly_fragment:01509
fragment_chain:1"
misc_feature
5269..13750
/note="assembly_fragment:00853
fragment_chain:1"
misc_feature
13851..17241
/note="assembly_fragment:01124
fragment_chain:1"
misc_feature
17342..25111
/note="assembly_fragment:01390
fragment_chain:1"
misc_feature
25212..33019
/note="assembly_fragment:00633
fragment_chain:1"
misc_feature
33120..50216
/note="assembly_fragment:00403
fragment_chain:1"
misc_feature
50317..62761
/note="assembly_fragment:01321
fragment_chain:2"
misc_feature
62862..84579
/note="assembly_fragment:01520
fragment_chain:2"
misc_feature
84680..98564
/note="assembly_fragment:00709
fragment_chain:2"
misc_feature
98665..101753
/note="assembly_fragment:00984
fragment_chain:2"
misc_feature
101854..114841
/note="assembly_fragment:01175
fragment_chain:3"
misc_feature
114942..122333
/note="assembly_fragment:01241
fragment_chain:3"
```

```
misc_feature 122434..130729
/note="assembly_fragment:00865
fragment_chain:3"
misc_feature 130830..135670
/note="assembly_fragment:00301
fragment_chain:3"
misc_feature 135771..141713
/note="assembly_fragment:01120
fragment_chain:3"
misc_feature 141814..151588
/note="assembly_fragment:01095
fragment_chain:3"
misc_feature 151689..176951
/note="assembly_fragment:00321
fragment_chain:3"
clone_end:sp6
vector_side:right"
BASE COUNT 53718 a 34261 c 33897 g 53373 t 1702 others
ORIGIN
Query Match 8.1%; Score 40.4; DB 2; Length 176951;
Best Local Similarity 53.1%; Pred. No. 3;
Matches 86; Conservative 0; Mismatches 76; Indels 0; Gaps 0;
QY 6 TGAACATAAAGCAGCACAATATTTTCAAGCTAGTCATTAATTTACTTGAATTCAT 65
D 57876 TTAATTAATATATACCTTCTTAAATATATGTCGAAAAAGCTTAATATGATATGAACATCAAT 57935
QY 66 TCGCATATTTGTGTACATTAATTAATTTAGATTGTTTCGTCTGCCAAGCCATCCCAATTA 125
D 57936 TTGCGAATAATCAATATCTATGAAATTTATATATATCAACACTATTAATAAAGCAATGTCATCA 57995
QY 126 TCCATCATTTTGGCAGCCAAATCTCTTATTTGGTGCATTCAT 167
D 57996 TCCATTAANAATTCATTCATTCACGTCCTCTCTTCTGTGTACAT 58037
RESULT 10
AC005683 74673 bp DNA linear PRI 02-OCT-2000
LOCUS Homo sapiens BAC clone CTB-62N11 from 7q22, complete sequence.
DEFINITION AC005683
ACCESSION AC005683.2 GI:4753225
VERSION HTG.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 74673)
AUTHORS Sulston,J.E. and Waterston,R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
JOURNAL 99063792
MEDLINE 9847074
PUBMED
REFERENCE 2 (bases 1 to 74673)
AUTHORS Threide,J., Wohlmann,P. and Modde,T.
The sequence of Homo sapiens BAC clone CTB-62N11
Unpublished
JOURNAL 3 (bases 1 to 74673)
REFERENCE 3 (bases 1 to 74673)
AUTHORS Waterston,R.H.
Direct Submission
Submitted (15-SEP-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
JOURNAL 4 (bases 1 to 74673)
AUTHORS Waterston,R.H.
Direct Submission
Submitted (05-MAY-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
JOURNAL 5 (bases 1 to 74673)
AUTHORS Waterston,R.
Direct Submission
TITLE
```

JOURNAL Submitted (02-OCT-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On May 5, 1999 this sequence version replaced gi:4204336.

COMMENT

----- Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc>

Contact: sapiens@watson.wustl.edu

----- Summary Statistics

Center project name: H_RG062N11

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send mailto:egreen@nhgri.nih.gov, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

Clone CTB-62N11 is from the first release of the human BAC library CTB-978SK-B. The library contains cloned DNA from the male fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). This clone is available from Research Genetics, Inc. (<http://www.resgen.com>).

VECTOR: pBel0BAC11

Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is CTB-15F3, 200 bp overlap; the clone sequenced to the right is CTB-36G2, 200 bp overlap. Actual start of this clone is at base position 195 of CTB-62N11; actual end is at 31845 of CTB-36G2.

FEATURES

Source

1..74673
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q22"
/clone="CTB-62N11"
/clone_1lb="CTB-978SK-B"
307..374
/rpt_family="L2"
648..678
/rpt_family="AT_rich"
2294..2602
/rpt_family="Alu"
2897..3017
/rpt_family="MIR"
3254..3582
/rpt_family="MERL-group"
3663..3982
/rpt_family="MERL-type"
4015..4273
/rpt_family="L2"
repeat_region
5163..5221

/rpt_family="L2"
5311..5708
/rpt_family="MaLR"
repeat_region
5709..6007
/rpt_family="Alu"
repeat_region
6008..6061
/rpt_family="MaLR"
repeat_region
6180..6629
/rpt_family="MaLR"
7642..7718
/rpt_family="MIR"
7975..8370
/rpt_family="MaLR"
repeat_region
9210..9251
/rpt_family="L2"
repeat_region
9293..9770
/rpt_family="L2"
repeat_region
9821..9936
/rpt_family="MERL-type"
9958..10234
/rpt_family="Alu"
10322..10483
/rpt_family="L2"
10487..10653
/rpt_family="MaLR"
10962..11619
/rpt_family="L1"
11742..11906
/rpt_family="MERL-type"
12278..12496
/rpt_family="L2"
12497..12786
/rpt_family="Alu"
12787..13190
/rpt_family="L2"
13288..13822
/rpt_family="L2"
13816..14069
/rpt_family="L2"
14148..14257
/rpt_family="MIR"
14421..14470
/rpt_family="G-rich"
14898..14918
/rpt_family="AT_rich"
15485..15775
/rpt_family="Alu"
16602..16809
/db_xref="GI:859023"
17856..18120
/db_xref="GI:859034"
complement(16263..18500)
/db_xref="GI:3660631"
20930..21613
/rpt_family="L2"
21614..21726
/rpt_family="MERL-type"
21727..22019
/rpt_family="L2"
22031..23210
/rpt_family="L1"
23225..23440
/rpt_family="L1"
24059..24101
/rpt_family="(CA)n"
24269..24297
/rpt_family="AT_rich"
24456..24708
/rpt_family="L2"
26522..26621
/rpt_family="L1"
30969..31166
/rpt_family="MIR"
repeat_region

Funding
Agency : Deutsche Forschungsgemeinschaft (DFG).
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
Location/Qualifiers
1. .55470
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/chromosome="2"
/map="3542416-3597884"
2458. .6183
/note="ORF_ID:dd_03152"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAM43736.1"
/db_xref="GI:21166119"
/translation="MEDEETRLNERRRIRREERKKOEEDADKEDQERKAREERY
RKEARTEKGEREAQOQKELEYIKRTEERORREERREERKEEKOQKEDEERK
QOERERRERREDEERKKIKELESRRKVLKELEDENSSQVDSILSKRRLIARO
KIESILASVDSNSQYOTOPTPIKIYSLILDELKEEVIDLASGGRRVADMSOP
ASISMAEYLVDIVGATIPASKPEPLNGTASLEKSLAMQASPEYLKIGCIOTAK
IITDPSNFIIDILPSMISNPPPIPLTMHLFTDIVCSYSRKPELLICNNKDTIL
PSLMQHSNSTEVDLLKMAEILEMTOSETSSSSGSAASFEFLTGSSSSSS
NNNNNSNNKKEPMTNDLTPYLILRINEISLENVIVDQLEIDNORIMFELYK
YKRSKVAHLGCGSFNAVLNRGFKLAVHPYEPFSSSMILCTIVDSVMTOSL
ITNSS
ELSPKATSELTITKDSOEFPYIRKLCIHLISINXANYSKIDREIWEIGILAKCI
DIFREYENNYIRHCIBITLPLIOLRCGSDDEPCHLDCSFVSKVNVILSGNN
SELQOINNSNN
QOQSEPVIGAPAPAMGSLSRGGKIRKRVSEFTDEDSSTADENNTSNTGVAYE
LAIDLSMFQPTPTPAVATTEOSSDKNSLENTVVSSEKRFATLTSTPISSSSS
SSTSTSTSTTTTTRKINCTSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS
LTPOEKAIKSOIKLKKERKKKKELAKLKEKESKSSSTSSSSSSSSSSSSSSSS
GSVLPSTPNKLKSSSSSPSSSPVLTSTTSTAATTPTTTTPTTTTPTTTTPTTTT
TWTTTTITSLSDSDTSNSDCSNDSVSEISKSLDNSOPTKSVENIESTIE
QSVVYVHNESSSNNSSNNTCNNSKINDSNTPTTSTTATKKGAVNHGHI
SLCIEITHIVNKSQGLRIDYDECCLFGQGMDEVVVKLQELISFRINRIGTVSMR
KGRGAPPTPEANVESFQNMANGVIPPTNSSKFF"

CDS

/note="ORF_ID:dd_03147"
/codon_start=1
/product="PROBABLE ATP-DEPENDENT RNA HELICASE"
/protein_id="AAM43738.1"
/db_xref="GI:21166121"
/translation="MAEDDGLQNLICLYLORPVKEINSAKEEAERYENISNNNFT
NNRAYREKTNKNN
IENNSNN
IESNNNEKDKYFAENKTKKDEFNRLPETIESNIYKESFSSPMWGLSITSETL
VRNLVGMKHEKPHHIOEASTPTILKGNDAVYKQSGKSLIPLIYVQGLTQART
RSDCYCYITPTRELSQIYEELQKLKPYVAVPGIIMGENRSARAKIRGINI
LVATPGRLLDHLQNTQSPDNINIKCLIDEDKLLDGFEDVTIINLDSKRRTK
FKRONILVSATLSGLISLASLSLSPVYIGLDSKVLKGENPQAAEKEMQLAPKL
DOFVEVESKRLTSLAFIRWTSNITIDGVAAGSSNSKAIYEFSSCSDVDH
HYMFSNMKMDKERGVKRTKKEOIKODKLIQHKQNSKIPTGNSDESDSDSDSD
SDSNSSEFDEKEIEKQIKETANSNRYKRTSVSPYPIFKLGEDDOQRTTFPD
FKNSPGLILTTDVSAKGLDPSVNMVVOVDPGSDTYDHIRIGRTARLQGGCSLLF
LLEPKRYIDHLAKFVNSYKEMVTTILOSLEPYSDQKLTKTSQSQLESQVHDLQL
FERLLYDSAKEMARCAIOSFLRSYATHKADVYSIFHISYHLGHVSKFALRETFP
ELNKSAGIKGAKAKKGGEDTLQOTADPFKMKYKSVNERSDGLINDOPORANFHE
SASGIVHNIYORHNRRNREADKNSNPKETKKRRLYDKKIESNKKFKK"
complement(join(10317. .10573,10659. .10671))
/note="ORF_ID:dd_01384"

CDS

/codon_start=1
/product="hypothetical protein"
/protein_id="AAM43739.1"
/db_xref="GI:21166122"
/translation="WTLIASISSIGNVKSITKSKNVRSSSSSSSSSSSSSSSSSS
GPLGANGVNLVGLIGTGVIVGSLVTFVGTINPILHPSGCGH"
join(12494. .12506,12587. .12843)
/note="ORF_ID:dd_01374"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAM43740.1"
/db_xref="GI:21166123"
/translation="WTLIASISSIGNVKSISKSNNFSSLSNSSLQSSNSIQCGCGGG
SPLIGVGNLVGVLVGTGIIIGTVGVTVNGVAGGLGSPNCGCH"
complement(join(13453. .13706,13780. .13792))
/note="ORF_ID:dd_01373"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAM43741.1"
/db_xref="GI:21166124"
/translation="WTLIASISSIGNVKSISKSNNFSSLSNSSLQSSNSIQCGCGGS
PLIGVGNLVGVLVGTGIIIGTVGVTVNGVAGGLGSPNCGCH"
join(14479. .14491,14572. .14825)
/note="ORF_ID:dd_01372"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAM43742.1"
/db_xref="GI:21166125"
/translation="WTLIASISSIGNVKSISKSNNFSSLSNSSLQSSNSIQCGCGGS
PLIGVGNLVGVLVGTGIIIGTVGVTVNGVAGGLGSPNCGCH"
complement(join(14976. .15348,15627. .15685,15765. .16167))
/note="ORF_ID:dd_01370"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAM43743.1"
/db_xref="GI:21166126"
/translation="MAVLTDSLTLNIGSMILLITPYILGPIINYYKKKTNPPTPA
LIVTGKSSIGGLALEYAKKRTNNSVLGTLGRNIEKLEIEICLDLGQVETESI
DVTGKASYDMLIFPNKYKIDILIANAGVETMLPELDETERILVTNTNVTGTLN
TVLWNPVLFESRGGGOLVMSITPPEYTMAGYSSSGYKSGFLIRNLASRGV
ISLITIPGVYPLVDSLDYKIDIPMLPTTPGKAKYIVDGISRDALITSCPVLVCS
HEFATIPPTLEADATNFINSIAKYPDKRPYTHSHHSRYSTNSKRENCSSSVKML
NSQIKITHNK"
complement(join(16864. .17173,17523. .17530))
/note="ORF_ID:dd_01367"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAM43744.1"
/db_xref="GI:21166127"
/translation="MLFSLKKNFTNLYLFNIKULPSISTIGNVKSISKSNNFSSLSN
SSLQSSNSIQCGCGGPLIGTVGNLVGVLVGTGIIIGTVGVTVNGVAGGLGSPNC
GCH"
join(17865. .17877,17959. .18215)
/note="ORF_ID:dd_01388"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAM43745.1"
/db_xref="GI:21166128"
/translation="WTLIASISSIGNVKSISKSNNFSSLSNSSLQSSNSIQCGCGGG
NSLIGTVGVLVGTGIIIGTVGVTVNGVAGGLGSPNCGCH"
complement(join(18637. .18888,18961. .18973))
/note="ORF_ID:dd_01385"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAM43746.1"
/db_xref="GI:21166129"
/translation="WTLIASISSIGNVKSISKSNNFSSLSNSSLQSSNSIQCGCGGG
SPLIGVGNLVGVLVGTGIIIGTVGVTVNGVAGGLGSPNCGCH"
join(19666. .19678,19759. .20015)
/note="ORF_ID:dd_01382"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAM43747.1"

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

```

/db_xref="GI:21166130"
/translation="MTLLAIISSIGNKKSIKSNFSSLSNLSQSNINIQCGCGGG
SPILGTGNIAGVGALVTGIIIVGTGVGVGAGLLSGNCCG"
complement(join(20170..20736,20806..20864,20943..21345))

```

```
/product="Putative Oxidoreductase"  
/protein_id="AAM43748.1"  
/db_xref="GI:21166131"
```

translating="MGWLDISLLIINIGSLIIFDIPGLGFLVYKKKKNFPPAA
IVTGASSGSGRGRLAETGAYARRKKNNLSLVGLGNRMNLEETQKCEGLGIVQVELES
DYDMDKELNDMLKPKKKYKIDILLIINACTIFILLPKELDFERLLIYANNTVIGTLN
TVLDPWPEIYFSSRSGGVVLAMSSITPFDIVMAQYSSSKYKIKFGLIARRSAGVG
VSVLTPEEITFPLVSDYKDKIDPEMLPIEDTVDAASKYIVDGLSRDAILISPLLYGIC
HFVATIPPTLRDATTNFSNLSLEKYSPKPTHYSHSQYSTNEKREDCCSSAAMNS
NSQOKREA"
21759..22802 |
/note="ORF_id:dd_01376"
/codon_start=1
/product="CYSTEINE-RICH, ACIDIC INTEGRAL MEMBRANE PROTEIN
PRECURSOR"
protein_id="AA043749.1"

Query Match	8.0%;	Score 40;	DB 2;	Length 55470;
Best Local Similarity	51.7%;	Pred. No. 3.8;		
Matches 91;	Conservative 0;	Mismatches 85;	Indels 0;	Gaps 0;

V

38 AAGTCATAAAATTTACTGCAITTCGAATTGCATTGTGGTACATAATTAGAATTGCGA 25935 AATTACAAGTAATTAATAATAATTAATTTAAATTCGAATATATATTAATTTAAATTTAA	97 25994
---	-------------

98 TTTCGTGTTTCCCAAGTCCACCAATATCCATCCCATTTGGCAGCCAATCTCTATTG 157
||| | | ||| | | ||| | | ||| | |
25995 TTTTATTTCACCAATTTATTATAATTAAATCAATATTAATCAAGTGAAAAATATAAATG 26054

158 GTGCATACATACATGAGTATCTTTCCACAAGAACACAGTTAGACCAAGTCCTTT 213
|| | | | | | | | | | | || | | | |
26055 GTAAAAAGTTTTTATCAAATTGATCCATCAACATCATCAATTTTAGAAAATAGTCATTTT 26110

RESULT 13			
C1194664/C			
PC1194664	150050	ba	1/1/2000
DATE		DATE	TIME 10.00

```

DEFINITION
Rattus norvegicus clone CH230-329N4, *** SEQUENCING IN PROGRESS
***, 49 unordered pieces.
AC119464
AC119466 3 CT.31746710
AC119468 3 CT.31746710

```

ORGANISM	SOURCE	WORDS
Elpharigata: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:		
Rattus norvegicus	Norway rat.	HTG; HTG5_PHASE1.

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 150950)

Aisodooks, S.L., Amarantunge, H.C., Are, J.R., Ayele, M., Banks, T.,
Barbaria, J., Benton, J., Bimaye, K., Blankenburg, K., Bonnin, D.,
Bouch, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
Buhay, C., Burch, P., Burlett, C., Burrall, K. I., Burd, N. C.

Carion, I. F., Carter, M., Cavazos, S. R., Cnacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C. D., Cox, C., Coyle, M. D., Dathorne, S. R., David, R., Davila, M. T., Davis, C., Davy-Carroll, I., Dedovich, D. A.

Deane, K.K., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraz, D., Flaga, N., Ford, I., Foster, P., Frantz, P.

Gadgil, M., Gao, J., Garcia, A., Gahleitner, I., Galzda, N., Gill, K., Gorrell, J. H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hatt, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgeson, A., Hommes, M., Holloway, C., Hollings, B.

nomis, F., Nowak, S., Rude, J., Rulyk, S., Hume, J., Jackson, L. E.,
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,

TITLE	Direct Submission	
JOURNAL	Unpublished	
REFERENCE	2 (bases 1 to 150950)	
AUTHORS	Morley, R.C.	
TITLE	Direct Submission	
JOURNAL	Submitted (27-Apr-2002)	Human Genome Sequencing Center, Department

REFERENCE
AUTHORS
3 (bases 1 to 150950)
MORLEY, K.C.

JOURNAL Submitted (18-Jul-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Jul 14, 2002 this sequence version replaced gi:20387289.

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hjsg.bcm.tmc.edu/>

```
----- Project Information -----
Center project name: GUXH
Center clone name: CH230-329N4
```

Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 1008 of reads
Assembly program: Phrap, version 0.990329

Consensus quality: 108058 bases at least Q30
Consensus quality: 111642 bases at least Q20

- * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
- * NOTE: This is a 'working draft' sequence. It currently
- * consists of 49 contigs. The true order of the pieces

- * arbitrary. Gaps between the contigs are represented as
- * runs of N, but the exact sizes of the gaps are unknown.
- * This record will be updated with the finished sequence.

* be preserved.
* 1 1015: contig of 1015 bp in length
* 1016 1115: gap of unknown length
* 1116 1115: gap of unknown length

*	2167	2266: gap of unknown length
*	2267	4049: contig of 1783 bp in length
*	4050	4149: gap of unknown length
*	4150	4244: gap of 1125 bp in length
*	5004	5004: gap of 1125 bp in length

*	5285	5384: gap of unknown length
*	5385	6410: contlg of 1026 bp in length
*	6411	6510: gap of unknown length
*	6711	7704: gap of 1014 bp in length

Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H.,
Lozano, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Meheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E.,
Massey, E., Mawhney, E., McLeod, M. P., Meador, M., Mel, G., Metzger, M.,
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokweto, S., Ogih, M., Okwuon, G.,
Ogunyeye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pichens, R., Primus, E., Pu, L., Quiles, M., Ren, Y.,
Rives, M., Rojas, A., Rojokan, I., Rolfe, M., Ruiz, S., Savary, G.,
Scherer, S., Scott, G., Shen, H., Shoostari, N., Sisson, I.,
Sodergren, E., Sonalike, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansley, J., Taylor, C., Taylor, T., Telford, B., Thomas, R., Thomas, S.,
Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S.,
Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G., and Gibbs, R.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

2 (phases 1 to 103138)
Worley, K.C.
Direct Submission
Submitted (19-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (phases 1 to 103138)
Worley, K.C.
Direct Submission
Submitted (17-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 13, 2002 this sequence version replaced gi:20303335.

COMMENT

Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

Project Information
Center project name: GSIM
Center clone name: CH230-163M8

Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 66968 bases at least Q40
Consensus quality: 71073 bases at least Q30
Consensus quality: 75068 bases at least Q20

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
NOTE: This is a 'working draft' sequence. It currently
consists of 51 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 1048: contig of 1048 bp in length
1049 1148: gap of unknown length
1149 2166: contig of 1018 bp in length
2167 2266: gap of unknown length
2267 3582: contig of 1316 bp in length
3583 3682: gap of unknown length
3683 5380: contig of 1698 bp in length
5381 5480: gap of unknown length
5481 6512: contig of 1032 bp in length
6513 6612: gap of unknown length
6613 7718: contig of 1106 bp in length
7719 7818: gap of unknown length
7819 9222: contig of 1404 bp in length

9223 9322: gap of unknown length
9323 10587: contig of 1264 bp in length
10587 10687: gap of unknown length
10687 11688: contig of 1003 bp in length
11688 11789: gap of unknown length
11789 13270: contig of 1381 bp in length
13270 13271: gap of unknown length
13271 14280: contig of 1019 bp in length
14280 14389: gap of unknown length
14389 16666: contig of 2277 bp in length
16666 16767: gap of unknown length
16767 18523: contig of 1757 bp in length
18523 18623: gap of unknown length
18623 20624: contig of 2000 bp in length
20624 20724: gap of unknown length
20724 21833: contig of 1110 bp in length
21833 21933: gap of unknown length
21933 23469: contig of 1536 bp in length
23469 23569: gap of unknown length
23569 25019: contig of 1450 bp in length
25019 25120: gap of unknown length
25120 26241: contig of 1122 bp in length
26241 26341: gap of unknown length
26341 27702: contig of 1361 bp in length
27702 27802: gap of unknown length
27802 28901: contig of 1099 bp in length
28901 29001: gap of unknown length
29001 30885: contig of 1884 bp in length
30885 30985: gap of unknown length
30985 32367: contig of 1382 bp in length
32367 32467: gap of unknown length
32467 33716: contig of 1249 bp in length
33716 33816: gap of unknown length
33816 35447: contig of 1631 bp in length
35447 35547: gap of unknown length
35547 36882: contig of 1335 bp in length
36882 36982: gap of unknown length
36982 38392: contig of 1410 bp in length
38392 38492: gap of unknown length
38492 40161: contig of 1669 bp in length
40161 40261: gap of unknown length
40261 41932: contig of 1671 bp in length
41932 42032: gap of unknown length
42032 43343: contig of 1311 bp in length
43343 43443: gap of unknown length
43443 44841: contig of 1398 bp in length
44841 44941: gap of unknown length
44941 46760: contig of 1819 bp in length
46760 46860: gap of unknown length
46860 47948: contig of 1088 bp in length
47948 48048: gap of unknown length
48048 49790: contig of 1742 bp in length
49790 49890: gap of unknown length
49890 51345: contig of 1455 bp in length
51345 51445: gap of unknown length
51445 53769: contig of 2324 bp in length
53769 53869: gap of unknown length
53869 55228: contig of 1360 bp in length
55228 55329: gap of unknown length
55329 58219: contig of 2790 bp in length
58219 58219: gap of unknown length
58219 60515: contig of 2296 bp in length
60515 60615: gap of unknown length
60615 62620: contig of 2005 bp in length
62620 62720: gap of unknown length
62720 63883: contig of 1163 bp in length
63883 63984: gap of unknown length
63984 65312: contig of 1329 bp in length
65312 65412: gap of unknown length
65412 68650: contig of 3238 bp in length
68650 68750: gap of unknown length
68750 71059: contig of 2309 bp in length
71059 71159: gap of unknown length

	repeat__region	/rpt_family="CGGGGgcn" 5835..5921 /rpl_family="(CGG)n" complement(8345..9535) /gene="OSJNBa0056617.17" /note="Predicted by genemark hm"	
gene	repeat__region	/complement(<8345..>9535) /gene="OSJNBa0056617.17" complement(8345..9535) /gene="OSJNBa0056617.17" /codon_start=1 /product="hypothetical protein"	
MRNA		/protein_id="AAG46168.1" /db_xref="GI:12039382" /translation="MADSHLDGVPMKRTVSSSSSLNIAFLSMVATNLISLHST RASPYPILLIPAAHQAOQDOLRLROLTAIRATYSQLNHRSPTPPPEPLLAYS RLAPLSASHPDLPLHFRRFNTYPEPCDDALSLAPLLRGCHPLPRRCRSSI SSHLNLNPDPSPPLPDSAVRWMPPGAKCTSPCLPSPGLPDLPAREARPLRACGP IDLYPOLLRILASISRAGRITGLDVGGGTTLAARKKRANMTYTITTNLCAPSE LAAGVGVPPLAQQRPVGDATMDIVRGHAANNRIPEALPEFWYADRVLRPG LLWDHFCRRPDLAAAYQPMRLRLGYKTLLKMAVDATPTPTAPAPGAKDHVYLAL LQKP"	
CDS	repeat__region	9206..9290 /rpt_family="(CGGGGgcn)" complement(9691..113329) /gene="OSJNBa0056617.2" /note="EST D47052.1 AU096093, AU075883 from this gene" complement(join(<9691..10099,10826..10955,11282..11459, 11659..11768,11903..12018,12376..12637,12766..13075, 13214..13329)) /gene="OSJNBa0056617.2" complement(join(9813..10099,10826..10955,11282..11459, 11659..11768,11903..12018,12376..12637,12766..13075, 13214..13329)) /gene="OSJNBa0056617.2" complement(join(12207..12227) /rpl_family="AT rich" complement(14187..16368) /gene="OSJNBa0056617.15" /note="highly similar to cholinephosphate cytidyltransferase GB:CBA70317 GI:1657382 (Pisum sativum); EST C26593 from this gene" complement(join(<14187..14282,14405..14515,14635..14703 14841..14996,15085..15127,15235..15330,15805..15918, 16059..16243)) /gene="OSJNBa0056617.15" /codon_start=1 /product="putative cholinephosphate cytidyltransferase"	
MRNA		repeat__region	repeat__region
CDS		repeat__region	repeat__region

[illegible]

Search completed: July 4, 2003, 16:40:42
Job time : 1513.67 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OW nucleic - protein search, using frame_plus.n2p model

Run on: July 4, 2003, 19:02:57 ; Search time 913 Seconds

(without alignments)
13505.226 Million cell updates/sec

File: US-10-083-053b-2
Perfected score: 52828

Sequence: 1 gatatgtaagaagcctca.....caattctgtaagaagtaa 29921

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 671580 segs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+np.model -DEV=xlh
-Q/cg12_1/USPTO.spool/US1008353.rnatc.03072003_093611.8089/app_query.fasta.1.30087
-DB=SPRMBL_21 -QFMT=fastan -SUFIX=isp -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blsum62 -TRANS=human40.cgi -LIST=45
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USRR=US1008353 @CGN_1.1.1549.ernat.03072003_093611.8089 -NCPU=6 -ICU=3
-NO_MMAP -LARGEIOBUF -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WANN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database 8

1: sprembl_21
2: sp_dacteria:
3: sp_fungi:
4: sp_human:
5: sp_invertebrate:
6: sp_mammal:
7: sp_mhc:
8: sp_organelle:
9: sp_phase:
10: sp_plant:
11: sp_rodent:
12: sp_virus:
13: sp Vertebrate:
14: sp Unclassified:
15: sp_virus:
16: sp_bacteriap:
17: sp_archaeap:

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match Length	DB ID	Description
1	5202	9.9	1275	4	000370 homo sapien

2	5195	9.9	1275	4	015604	015604 homo sapien
3	5190	9.9	1275	4	000360	000360 homo sapien
4	5188	9.9	1275	4	09Y5K0	09Y5K0 homo sapien
5	5188	9.9	1275	4	09UN80	09UN80 homo sapien
6	5186	9.9	1275	4	08TE30	08TE30 homo sapien
7	5182	9.9	1275	4	000366	000366 homo sapien
8	5182	9.9	1275	4	000375	000375 homo sapien
9	5176	9.9	1275	4	015606	015606 homo sapien
10	5175	9.9	1275	4	000363	000363 homo sapien
11	5169	9.9	1275	4	012881	012881 homo sapien
12	5168	9.9	1275	4	000378	000378 homo sapien
13	5163	9.9	1275	4	000362	000362 homo sapien
14	5150	9.8	1275	4	000368	000368 homo sapien
15	5133	9.8	1275	4	000372	000372 homo sapien
16	3605.5	6.9	1281	11	09QU12	09QU12 mus musculus
17	3605.5	6.9	1281	11	054850	054850 mus musculus
18	3600.5	6.9	1281	11	088913	088913 mus musculus
19	3597	6.9	1300	11	P97692	P97692 rattus norv
20	3595.5	6.9	1281	11	088915	088915 mus musculus
21	3594.5	6.9	1281	11	09QWY3	09QWY3 mus musculus
22	3590.5	6.9	1300	11	008906	008906 mus musculus
23	3589.5	6.9	1281	11	09QWY0	09QWY0 mus musculus
24	3588.5	6.9	1281	11	088914	088914 mus musculus
25	3559	6.8	1275	6	062658	062658 canis fami
26	3558.5	6.8	1281	11	091289	091289 mus musculus
27	3557.5	6.8	1281	11	060713	060713 mus musculus
28	3557.5	6.8	1281	11	091288	091288 mus musculus
29	3533.5	6.8	1300	11	061785	061785 mus musculus
30	3511.5	6.7	1252	11	09QWY2	09QWY2 mus musculus
31	3366.5	6.4	1219	11	09QWY3	09QWY3 mus musculus
32	2883.5	5.5	712	4	014754	014754 homo sapien
33	2653.5	5.1	641	4	014288	014288 homo sapien
34	2284.5	4.4	573	4	000549	000549 homo sapien
35	2237	4.3	611	4	014288	014288 homo sapien
36	2234	4.3	1275	4	000362	000362 homo sapien
37	2234	4.3	1275	4	000363	000363 homo sapien
38	2230	4.3	1275	4	000366	000366 homo sapien
39	2229	4.3	1275	4	015604	015604 homo sapien
40	2229	4.3	1275	4	09UN80	09UN80 homo sapien
41	2229	4.3	1275	4	015606	015606 homo sapien
42	2226	4.2	1275	4	09Y5K0	09Y5K0 homo sapien
43	2225	4.2	1275	4	000378	000378 homo sapien
44	2224	4.2	1275	4	000370	000370 homo sapien
45	2222	4.2	1275	4	000375	000375 homo sapien

ALIGNMENTS

RESULT 1
ID 000370 PRELIMINARY; PRT; 1275 AA.
AC 000370;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative p150.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=97285120; PubMed=9140393;
RX Sasanian D.M., Dombroski B.A., Moran J.V., Kimberland M.L., Naas T.P.,
RA Deberardinis R.J., Gabriel A., Swergold G.D., Kazazian H.H. Jr.;
RT "Many human l1 elements are capable of retrotransposition";
RL Nat. Genet. 16:37-43(1997).
DR EMBL: U93569; AAC51271.1; -;
DR InterPro: IPR005135; Exo_endo_phos.
DR InterPro: IPR000477; RYISe.
DR Pfam: PF03372; Exo_endo_phos; 1.
DR Pfam: PF00078; Ivt; 1.
KW RNA-directed DNA polymerase.

SEQ SEQUENCE 1275 AA; 149011 MW; 588703688E7129FF CRC64;
Alignment Scores:
Pred. No.: 0 Length: 1275
Score: 5202.00 Matches: 1059
Percent Similarity: 79.18% Conservative: 78
Best Local Similarity: 73.75% Mismatches: 133
Query Match: 9.94% Indels: 173
DB: 4 Gaps: 7
US-10-083-853b-2 (1-29921) x 000370 (1-1275)
QY 3428 ATGGCAGGATCAATTCACACATTAATATATTAACCTTAATGTAAGGGCTAAATGCC 3487
DB 1 MetThrGlySerAsnSerHisIleThrIleLeuThrLeuAsnValAsnGlyLeuAsnSer 20
QY 3488 CCAATTAAAGACACAGACTGGCAAAATGGATAAGGTCAAGACCCCATAGTGTGT 3547
DB 21 ProlIleYsArgHisArgLeuAlaSerTrpIleYsSerGlnAspProSerValCysCys 40
QY 3548 ATTCCAGAGGCCCATCTCACATGAAAAAGACACACATGAGCTCAAAATTAAGGATGGAG 3607
DB 41 IleGlnGluThrHisIleuThrCysArgAspThrHisArgLeuYsIleYsGlyTrpArg 60
QY 3608 AAGATTTCACACTAAATGGAACAAACAAAAAAGACAGGGGTGCATCTAGTCTCT 3667
DB 61 LysIleTyrGlnIleAlaAsnGlyLysGln---LysLysAlaGlyValAlaIleLeuValSer 79
QY 3668 GATFAAAACAGACTTTAAACCAACAAAGATCAAAAGACAAAGGCCATTACATATG 3727
DB 80 AspLysThrAspPheLysProThrLysIleLysArgSpLysGluGlyHisTyrIleMet 99
QY 3728 GTAAGGCAATCAATGGAACAAAGAGCTAACTATCTTAATATATGACGCCAATATCA 3787
DB 100 ValIleGlySerIleGlnGlnGluGluLeuThrIleLeuAsnIleTyrAlaProAsnThr 119
QY 3788 GAGGACCCAGATTCATTAAGCAAGTCTTAGAGACCTTACAAAGACATTGACTCCAC 3847
DB 120 GlyAlaProArgPheIleLysGlnValLeuSerAspLeuGlnArgAspLeuAspSerHis 139
QY 3848 ACATATATAGTGGAGCTCTTAATATATATAGACACTTTACACCCCATGCCAATATTA 3907
DB 140 ThrLeuIleMetLys-----AspPheAsnThrProLeuSerIleLeu 153
QY 3908 GGCAGATCAATGACAGAAATTAACAGGATATCCAGAGATTGACCTGACCTGGAC 3967
DB 154 AspArgSerThrArgGlnLysValAsnLysAspThrGlnGluLeuAsnSerAlaLeuHis 173
QY 3968 CAAGCGACCTTAATATATCTACAGACTCCACCCCAATCAACAGATATACACTC 4027
DB 174 GlnThrAspLeuIleAspIleTyrArgThrLeuHisProLysSerThrGluTyrThrPhe 193
QY 4028 TTCTCGACATCACTATACACCTATTTAATATGACCATATTTTAAGTAACAACTC 4087
DB 194 PheSerAlaProHisHisThrTyrSerLysIleAspHisIleValGlySerLysAlaLeu 213
QY 4088 CTCAGCAATTCGAAAAAGAACAGAAATCTTAACAAAGTCTCTCAGACTACAGTGCATC 4147
DB 214 LeuSerLysCysLysArgThrGluIleIleThrAsnTyrLeuSerAspHisSerAlaIle 233
QY 4148 TATTTGAATCTGAATTAAGAACTCCTCAAAATCACAACATCACTAGTGAAGTGAAC 4207
DB 234 LysLeuGlnLeuArgIleLysAsnLeuThrGlnSerAspSerThrThrTrpLysLeuAsn 253
QY 4208 AACCTCTCTGTAATGACTACTGCTGTAATTAACAAATGAAGGCAAAATTAAGATGTC 4267
DB 254 AsnLeuLeuLeuAsnAspLysTrpValHisAsnGluMetCysAlaGluIleLysMetPhe 273
QY 4268 TTGGAACCAATGAGAACAAAGACACATATACCAATCTCTGGGCAATATTATTAAGA 4327
DB 274 PheGluThrAsnGlnLysAspThrTrpTyrGlnAsnLeuThrPsrAlaPheLysAla 293
QY 4328 GTGTGTAGAGGCAAAATTTATAGCACTAGATGCTTACAGAGAAAGACGAAATATCTAAA 4387

DB 294 ValCysArgGlyLysPheIleAlaLeuAsnAlaTyrLysArgLysGlnGluArgSerLys 313
QY 4388 ATGACACCCCTTAACATCACATTAATAAAGACTGAGAGAGAAAGCAACAAATTCAAAA 4447
DB 314 IleAspThrLeuThrSerGlnIleLysGlnLeuGlnLysGlnIleGlnThrHisSerLys 333
QY 4448 GCTAGCAGAAAGACAGAAATTAAGTACAGACAGACAACTGAAAGAGATAGACACAA 4507
DB 334 AlaSerArgArgGlnGluIleThrLysIleArgAlaGluLeuLysGluIleGluThrGln 353
QY 4508 AAAGCCCTTAATTAATCAATGATACAGAGACCTGGTTTGGAAAGATCGCAAAAT 4567
DB 354 LysThrLeuGlnLys-IleAsnGluSerArgSerTrpPheGlnAlaGlyIleAsnLysIle 373
QY 4568 ----AGACACTAGACAGACTAATTAAGAAAGAAAGAGAGAAATCAAGACATCAAT 4623
DB 373 AspArgProLeuAlaArgLeuIleLysLysLysArgLysLysAsnGlnIleAspThrIle 393
QY 4624 AAAAATGATTAAGGGGATATCACACCCAGATCCACAGAAATCAAACTATTATCAGAG 4683
DB 393 LysAsnAspLysGlyAspIleThrThrAspProThrGlnIleGlnThrThrIleArgGly 413
QY 4684 ATATTTAACACCTCTATGCAATTAATACTAGAAATCTAGAAATGAGATTAATCTCT 4743
DB 413 GtyTyrLysHisLeuTyrAlaAsnLysLeuGlnLysLeuGlnIleMetAspThrPheLe 433
QY 4744 GACACATATATGTAGCTGTATGACCTTGGGGGACAGAACAAAGGGGTGAATGCAGAA 4803
DB 433 uasp----- 434
QY 4804 ATAAAGACAAAGACAAAGAGATGTGTGGAAAGTAGGGGTACAGGGGCAACTTGCCT 4863
DB 434 ----- 434
QY 4864 AATGCAAGAGGCCCTGAGCTTTACACACACCTCTGATTTATTAGCAAAAGATAGC 4923
DB 434 ----- 434
QY 4924 GAGAGGTGATTGGAAGAAAGAGTCACTGTTAGTCCAGAGTAGGCTGCAGACTGC 4983
DB 434 ----- 434
QY 4984 ATTCTCAAAACAATAGCTCTAGATGTCCAGTAGATTAACCTCAAGACCACTGCCAG 5043
DB 434 ----- 434
QY 5044 GAGTGAATGGCCCTGACAAACCTTGAAGGACAGACAGAAAGTAAGTTGCCACATTCT 5103
DB 434 ----- 434
QY 5104 GTATTCAGATTAACAGTTTGTCTTGTGATCAAGTAGTCCAGTGAATGCTGAGTTGG 5163
DB 434 ----- 434
QY 5164 TCATGATCCCTTTGGCTTTTGGCTCCAAAAACATACACCTCTCAAGACTAAACCA 5223
DB 435 ----- 435
QY 5224 GGAAGAGTCAATCCCTGAATATACCACTACCAATCTTAATATGAAGCACTATTA 5283
DB 443 nGluGluValGluSerLeuAsnArgProIleThrGlySerGluIleValAlaIleLeuAs 463
QY 5284 TAGCTTACCAACCAAAAAAGTCCAGAGACAGAGGATTCACAGCCAAATTTACACAG 5343
DB 463 nSerLeuProThrLysLysSerProGlyProAspGlyPheThrAlaGluPheTyrGlnArg 483
QY 5344 GTACAAAGAGAAGCTGTACTATTCTCTTGAACATATTCACAAAATAGAA---AATGG 5400
DB 483 gTyrLysGlnGluLeuValProPheLeuLeuLysLeuPheGlnSerIleGluLysGln-G 503
QY 5401 GAATCTCTCCCTAATCTATTATTTACAGAGCCAGCATCATCTGTGATACCAAACTAGACG 5460
DB 5401 ----- 5401

Dd 503 lylleleuProasnSerPheTyrGlnAlaSerIleIleleuIleProLysProGlyArgA 523
Qy 5461 ACACAAACAAAGAGAAATTTGAGGCCATATCCCGATGAGCATTTGATGTGAAATCC 5520
Dd 523 sphrThrLysLysGluAsnPheArgProIleSerLeuMetAsnIleAspAlaLysIleL 543
Qy 5521 TCATATAAATACTGGCAACCAATCCAGACGACATCAAAAGCTTATCTACCATGATC 5580
Dd 543 eunsLysIleleuAlaAsnArgIleGlnGlnHisIleLysLysLeuIleHisIleAspG 563
Qy 5581 AAGTGGCGTCATCCCTGGGATGAGCGCTGTTCAAAATATGCAATCAATAATGTAG 5640
Dd 563 lvalGlyPheIleProGlyMetGlnLysPheAsnIleArgLysSerIleAsnValI 583
Qy 5641 GCCATCATAAACAGAACCAATGACAAAACACATGATTCATCTCAATGATGCAGAAA 5700
Dd 583 lcglnHisIleAsnArgAlaLysAspLysAsnHisValIleIleSerIleAspAlaLul 603
Qy 5701 AGGCCCTTTGCAAAATTCACACGCCCTTCATGCTAAAAATTCAGTAACTAGTATCG 5760
Dd 603 lysAlaPheAspLysIleGlnGlnProPheMetLeuLysThrLeuAsnLysLeuclYlLea 623
Qy 5761 ATGGAATGATCTCAAAATATAAGAGCTTTATAC-AAACCCACAGCCCAATATCTAC 5819
Dd 623 spgIleTyrLysLysIleIleArgAlaIleTyrAspLysProThrAlaAsnIleIleL 643
Qy 5820 TGAATGGGCAAAACCTGAGACATTCCTTGAGAACTGGCACAGACAGATGCCCTC 5879
Dd 643 eunsngLysLysLysGlnAlaPheProLeuLysThrGlyThrArgGlnGlyCysProL 663
Qy 5880 TCTCACACATCTCATTCACAACTATGTAAGTTCTGGCCAGGCAATCAGGCATAGA 5939
Dd 663 euserProleuLeuPheAsnIleValLeuGlnValleuAlaArgAlaIleArgGlnLul 683
Qy 5940 AAGAATAAAGGCTATTCATATAGAAAGAGAGAGATCATTTGCTCTGTTGCAGATG 5999
Dd 683 lysGlnIleLysGlyIleGlnLeuLysGlnLysLysLysLysLysLysLysLysLysLys 703
Qy 6000 ACATGTTTGTATTTTGAAGAACCCCATGCTCAGGCCCAAAACCTCTTAAGCTGTAA 6059
Dd 703 spmetIleValIleTyrLeuGlnAsnProIleValSerAlaGlnAsnLeuLysLeuLies 723
Qy 6060 GCAACTTCAGCAAAAGCTCAGAGACACAAATCAATGTGCAAAATCAACAGATTTCTAT 6119
Dd 723 ersnPheSerLysValSerGlyTyrLysIleAsnValGlnLysSerGlnAlaPheLeuT 743
Qy 6120 ACCGCATATATAGCAAAACAGAGAGCCAAATCATGAGTCACTCTTACAACTTCTTA 6179
Dd 743 lrsnAsnAsnArgGlnThrGlnSerGlnIleMetGlyLysLysLysLysLysLysLysLys 763
Qy 6180 CAAGAGAAATTAATACCTAGGAATACATTCACAAAGGACAGTGAAGTCTTCAAG 6239
Dd 763 erLysArgIleLysTyrLeuGlyIleGlnLeuThrArgAspValLysAspLeuPheLysG 783
Qy 6240 AGAAGTCAAAACAGCTGATCAAGAAATTAAGAGAGACACAAACAAATGAAAAACATTC 6299
Dd 783 lunsnLysLysProleuLeuLysGlnLysLysLysLysLysLysLysLysLysLysL 803
Qy 6300 CAGTCTCAGAGATAGTAGAATCAT-----GAAATGCCATATCTGCCAAAGTAAT 6352
Dd 803 rocLysSerTyrValGlyArgIleAsnIleValIleLysMetAlaIleLeuProLysValIleT 823
Qy 6353 ATGATATGATGCTGACCCCATCAGCTATCATTTGATCTTCTACAGATTTGAAAAA 6412
Dd 823 lrsnArgPheAsnAlaIleProIleLysLeuProMetThrPhePheThrGlnLeuGlnLysT 843
Qy 6413 CAACCTTAATTTTCATATGGAACCAAAAAAGAGCCACAGACAGCAACAGCAATCTTAAGC 6472
Dd 843 hrThrLeuLysPheIleTyrPasnGln-LysArgAlaArgIleAlaLysSerIleLeuSer 862
Qy 6473 AAAAAAGCAAAAGCTGAGATCATGCTTACCTGACTTAAACTATATCTTAAGGCTACA 6532
Dd 863 GlnLysAsnLysAlaGlyIleThrLeuProAspPheLysLeuTyrLysAlaThr 882

Qy 6533 GTAACCAAACTGCATGCTACTGTACCAAAACAGATATATACCAATGAGACAGACA 6592
Dd 883 ValThrLysThrAlaTyrTyrGlnAsnArgAspIleAspGlnThrPasnArgThr 902
Qy 6593 GAGACCTCAGAAATTAAC-ACATGCAATCTACATCATCTGATCTTTGCAAACTGACAAA 6651
Dd 903 GluProSerGlnIleMetProHisIleTyrAsnTyrLeuIlePheAspLysProGlnLys 922
Qy 6652 AACAGCAATGAAAAAGATTCCTCTTATTAATATGATGGTTGGAAAACTGGCTAGCC 6711
Dd 923 AsnLysGlnThrPglLysAspSerLeuAsnLysTyrPcysTyrPglAsnTyrPleuAla 942
Qy 6712 ATATGCAAGATGAAAGTGAATGATCCCTGCTTACACCTTATCAAAAGTAACTACAGA 6771
Dd 943 llecysArgLysLeuLysLeuAspProPheLeuThrProTyrThrLysIleAsnSerArg 982
Qy 6772 TGAATTAAGACTTAATATATAGACATTAACCAATAAAACCA-GAAGAAAACCTAGGC 6830
Dd 963 TrpIleLysAspLeuAsnValLysProLysThrIleLysThrLeuGlnGlnAsnLeuGly 982
Qy 6831 AATACCATTCAGGATATGACATGGGCAAGACTTCATGACTTAAACCAAAAGCAATG 6890
Dd 983 lIleThrIleGlnAspIleGlyValGlyLysAspPheMetSerLysThrProLysAlaMet 1002
Qy 6891 GCAACAAAGCCAAATAGACAAAGTGGATCTGATTAATCTATAGACTCTGCACAGCA 6950
Dd 1003 AlaThrLysAspLysIleAspLysTyrPasnLeuIleLysLeuLysSerPheCysThrAla 1022
Qy 6951 AAAAAAACTGTCATCAGACTGAGCAAGCAACCACTACAGAAATGGGAGAAAAATTTTGCAT 7010
Dd 1023 -LysGlnThrThrIleArgValAsnArgGlnProThrThrProLysIlePheAlaTh 1042
Qy 7011 CTATGATCTGACAAAGCTAATATCCAGAGATCTACGAAAGTAACTTAACAAATTTTCA 7070
Dd 1042 rTyrSerSerAspLysGlyLeuLeuIleSerArgIleTyrAsnGlnLeuLysGlnIleTyrLys 1062
Qy 7071 GAAAAA-----ACACCCCGCTCAAAATATGGGCAAAAGATATGACACACTTCTCAAA 7126
Dd 1062 slsLysLysThrAsnAsnProIleLysLysTyrPalaLysAspMetAsnArgIlePheSerLys 1082
Qy 7127 AGAAGCATTTATGACACCAACCAACATATGAAAAAAACCTCATCATCATTTGCTGTAG 7186
Dd 1082 sGlnAspIleTyrAlaLysLysHisMetLysLysCysSerSerSerLeuAlaIleArg 1102
Qy 7187 AGAAATGCAAAACAAACCAAGTGCATACCATCTCATCTGATTTGAATGGTGATCAC 7246
Dd 1102 gGlnMetGlnIleLysThrThrMetArgTyrHisLeuThrProValArgMetAlaIleIle 1122
Qy 7247 TAAAAAGTCAAGAAACCAAAATGCTGAGAGATGGAGAAATAGAACACTTTTCCA 7306
Dd 1122 elysLysSerGlyAsnAsnArgCysTyrPalaGlyCysGlyGlnIleLysThrLeuValHl 1142
Qy 7307 CTGTGGTGGGAATGTAATTAATTAAGTCAACCATTTGTGAGACAGGTGAGATTTCTTAA 7366
Dd 1142 scYsTyrTyrPasnCysLysLeuValGlnProleuTyrLysSerValTyrPrgPheLeuT 1162
Qy 7367 GGATCTGAGACCAAGAAATATCATTTTGACCCAGAAATCCCATTTACTGAGTATATACCAA 7426
Dd 1162 gasPleuGlnIleLysLysThrThrMetArgTyrHisLeuThrProValArgMetAlaIleIle 1182
Qy 7427 GGAATTAATCACTTATTAATTAAGACAGATGACATGATGTTTATTCACAGACATGAT 7486
Dd 1182 sasPtyrLysSerCysCysTyrLysAspThrCysThrArgMetPheIleAlaLeuPn 1202
Qy 7487 CACAATAGCAAAAGCTTGAACCAACCAACCAATGTCCATCAGTATGATGATGATTAAGAA 7546
Dd 1202 eThrIleAlaLysThrTyrPasnGlnProAsnGlyCysProThrMetIleAspTyrLysLys 1222
Qy 7547 AACATGGCAATATATACCATGAAATATCTATGAGGCATTAAG-GATGAGTTCATGTC 7605
Dd 1222 sketTyrPasnIleLysThrMetGlnTyrTyrAlaAlaIleLysAsnAspIlePheIleSe 1242

QY 4984 ATTCTCAAAACAATGAGCTCTAGATGTCCTCCAGTAGATACTCAAGAGACCAGTCCAGG 5043
Db 434 -----
QY 5044 GAGTATGAGCCCTCAGCAAACTTCTAGGGCAGGCACAGAGTAAGTTGGCCACATTCT 5103
Db 434 -----
QY 5104 GTATTACAGATAAAGATTGCTGTTTGATCAAGTAGCCTCCAGTGGATGCTGAGTTGG 5163
Db 434 -----
QY 5164 TCATGATCCCTTTGGCCCTTTTGGCTCCCAAAACATACACCTCTCAGACTAACA 5223
Db 435 -----
QY 5224 GGAAGAGTCAAAATCCCTGCAATATACCAATCAAGTTCTAAATTTGAAGCAGTAATTGA 5283
Db 443 nglugluvalgluserleuasnarprollethriglsersgluilevalalalleas 463
QY 5284 TGGCTTACCAACCAAAAAAGTCCAGACAGCAGATTCACAGCCTTCTACAGAG 5343
Db 463 nserleuprotlrlylserserproglproaspglypnetlrAlaGluphetrlyclnAr 483
QY 5344 GTACAAAGAGAGCTGTACTATCTCTTGAACACTATCCAAAAATAGAA---AATGG 5400
Db 483 gtyrmetglugluvalprophetleuulysleupheginserilleglulyslu-G 503
QY 5401 GAATCCCTCCCTACTCATTTTACGAGGCCAGCATATCCGTATACCAAAACCTAGCAGTG 5460
Db 503 lytleuprosanserphetrlygluAlaserilleleulleprolysproglyArGa 523
QY 5461 ACACAACAAAAAGGAATTTTCAGGCCCATATCCCTGATGAACATGTATGGAAAAATCC 5520
Db 523 sptlrthrltlysltysgluasnpheargprolleserleuemasnilleaspaAlaLysleL 543
QY 5521 TCAATAAAATACTGGCAAAACCAAAATCCAGCACATCAAAAAGTTATCTACCATATC 5580
Db 543 euasnllylleleuAlaasnarglleglnglnhlsrlyslsleuilehshlsaspg 563
QY 5581 AAGTTGGCGTCATCCCTGGGATGCAAGCGTGTCAAAATATGCAATCAATCAATATGAG 5640
Db 563 llnvalgllypheitlleproglmetglnglytrpPhesnllearglysserilleasvalil 583
QY 5641 GCCATCACATAACGAACCATGACAAACCAACCATGATTTATTCATATAGATGCGAAA 5700
Db 583 leglnhlsrlyleuasnarAlaasnaspLysasnhsrmetilleleiserilleaspaAlaGlul 603
QY 5701 AGCCCTTTGTCAAAATTCACAGCCCTTCATGCTAAAATTTCTCAGTAACATAGATATCG 5760
Db 603 ysAlapheaspLyslleglnglnprophemetleuulysrhrleuasnllyslenglylIla 623
QY 5761 AAGCAATGTATCTCAAAATATATAGAGCTATTTATAC-AAACCCACAGCCCAATATCATAC 5819
Db 623 spgllythrlyrphesyllelleargAlaalletraspLysprothrAlaasnilleleL 643
QY 5820 TGAATGGGCAAAAACGGAAGCATGCCCTTGAGAACTGGCACAGAGAGAGATGCCGTC 5879
Db 643 euasnglyslnlyslsleugluAlapheproleuulysrhrlythrarglnglyCysproL 663
QY 5880 TCTCACACTCTATTCAGATACTATTTGGAAGTTCTGGCCAGGGCAATCAGACATAGA 5939
Db 663 euserProleuulysleuasnllevalleugluvalleuAlarAlaIlearglngluL 683
QY 5940 AAGAAATAAAGGATTCATAATAGAAAGAGAGAGAGACTCATATTTGCTCTGTTGCAAGATG 5999
Db 683 yslgluilelysllylleglnglnlyslgluAlaLysleuSerleupheAlaaspa 703
QY 6000 ACATGTTTATATTAGAAAACCCCATGCTCAGGCGCAAAAACCTCTTAAGCTGATGA 6059
Db 703 spmetllevallytrleugluasnprolilevalserAlaGlulnsnleuulysleuilles 723

QY 6060 GCAACTTCAGCAAAAGTCTCAGACACAAAATCAATGTGCAAAAATCACAAGCATTCCTAT 6119
Db 723 eRAsnpheserlyslvalsergllytrlyslleasnvalglulnysserglinalaIleuL 743
QY 6120 AGCCCAATATATAGCAAAACAGAGAGCCAAATCATAGTGAACCTTCATTCACAATTGCTA 6179
Db 743 yThrAsnasnarglnthrArgluserglnllemetclylgluLeuproPhevalilleas 763
QY 6180 CAAGAGATTAATAATACCTAGGAATPACACTTACAAAGGACGTAAGCAACTTCACAG 6239
Db 763 erlysnrgllyleuyllylleuglylleglulnleuulrargaspAlaLysaspleuPheylsG 783
QY 6240 AGAATPACAACCACTGATCAAGGAATTAAGAGAGACACAACAATGGAATAATTC 6299
Db 783 luanstlyrsproleuulysgluilelysgluinsphrlysnlystrplysaNlleP 803
QY 6300 CATGCTCAGCATAGTAAGATCAT-----GAAATGCCATCTGCCCAAGTAAT 6352
Db 803 rocyserrtrpAlaIarglileasnillevallysmetalalleuProllyvalilleT 823
QY 6353 ATAGATTGCTGCTACCCCATCAAGCTACCATGTGACTTCTTCACAGAAATGGAATAA 6412
Db 823 yThrAsnasnAlaIleprollylsleuprometthrPhePnetrgluLeuGlulysT 843
QY 6413 CAACTTAATTTTCATATGGAACCAAAAAAGGCCACAGAGCCCAAGCAATCTTAAC 6472
Db 843 hrThrleuulysrheiletrrpaNgln-LysArgAlaArglilleAlaLysserilleuSer 862
QY 6473 AAAAGACAAAGCTGAGAGTATCATGCTGACTTAAACATATATAGCTTACA 6532
Db 863 GlnLysasnlyAlaGllyllylethrleuproaspPheleuyltyrlysalatrhr 882
QY 6533 GTAAACCAAAAGCTGAGTACGTACCAAAAACAGTATATAGCAAAATGGAACGAACA 6592
Db 883 valthrlyslhrAlatrpyrrpyrrgluasnaArgaspLiespdlntrrpsnarGthr 902
QY 6593 GAGACTCAGAAATTAAC-ACAGCAATCTCATCATCATATCTTTGACAAACCTGCAAA 6651
Db 903 GluproserclulilemetProhlsrlyletrAsnlyleuilelphesaspLysproglulys 922
QY 6652 AACCAAGCAATGCAAAAAGATTCCTATTTAATAATGGTGTGGAAAACTGGCTAGCC 6711
Db 923 AsnlysglntrpGllyLysaspsrleupheasnlystrpCytrpLulnsnrpleuAla 942
QY 6712 ATATGCAAGAGCTGAAACTGGATCCCTTCCTTACACCTTATACAAAATTAACACAGA 6771
Db 943 lleCysArglyleuulysleuasnpProphetleuthrProtyThrlyslleasnserArg 962
QY 6772 TGAATTAAGACTTAATATATAGACATAAACCATAAACCACCA-GAAGAAAACCTAGGC 6830
Db 963 trpIlelyAspleuasnlvallyspLysrhrillelystrhrleugluLulnsleugly 982
QY 6831 AATACATTCAGATATGACATGGGCAAAAGCTTCATACCTAAAACCAAAACGATG 6890
Db 983 lIethrilleglAspIlellyAlaGllyLysasprhemeserlystrProlysAlaMet 1002
QY 6891 GCAACAAAAGCCCAAAATATAGCAAGTGGATCGATTAACCTATAGAGCTTGCACAGCA 6950
Db 1003 AlatrhrLysAspLyslileasplystrpaspleuilelyslserPheylstrAla 1022
QY 6951 AAAAAAACTGTATCAGATGAGTGAACAGCAACCTACAGATGGAGAAAAATTTGGCAAT 7010
Db 1023 -LysgluThrthrIleargvalAsnarglInProthrtrtrpGlulysillePheAlaThr 1042
QY 7011 CTATGATCTGACAAAAGGCTAATATCCAGAGATCTACGAAGAATTAACAATTTACAA 7070
Db 1042 rtyrserSerAspLysgllyleuileserArgliletrArglngluLeuLysgluIlelyrly 1062
QY 7071 GAAAAA-----AACACCCCGCTCAAAATATAGGCAAAAGATATAGCAGACACTTGTCAA 7126
Db 1062 sltyslystrhrAsnasnprolilelyslstrpAlaLysaspmetasnArghlshPheSerly 1082
QY 7127 AGAAGCATTTATGACGCCAACAAACATATGAAAAAACCTCATCATCTATGCTGCTAG 7186

```
Db 1082 sgluaplleleryralaalalyslyshismetlyshscysseiserleualallear 1102
Oy 7187 AGAATGCAAAAACCAACAGATGACATACCATCTCGTAGTAAAGTGTATAC 7246
Db 1102 gglumetclnlelystrthrmetyrlythshleuthrprovalaigmatalailell 1122
Oy 7247 TAAAAAGTCAGAAACAACAATGCTGAGAGATGTGAGAAATAGAACATTTTCCA 7306
Db 1122 elyslysserclysnasnancgystprargglycysglglnlleglythrleuenuh 1142
Oy 7307 CTGTGCGGGAATCTAATTAATGATTCACCATCTGGAGAACAGTGTGAGATTCCTTAA 7366
Db 1142 scystprtrpaspkyslyslleualglnproleuthrlysserValtrpargmleuar 1162
Oy 7367 GGATCTAGAACCCAGAAATATCATTTGACCCAGCAATCCATTAAGTAGTATACCCAAA 7426
Db 1162 gaspleucluleuenglulieprophesaprhoalallerleuenugllyletryprohs 1182
Oy 7427 GGAATATAATCATTCATTTATTAAGACACATGCACATATGTTTATGACAGCTGAT 7486
Db 1182 nglutrylssercyslytrylsaspthrCysThrargmetrheilealaaleuph 1202
Oy 7487 CACATAGCAAGATTTGGAACCAACCAATGTCATCAGATGATAGACTGGATAAAGAA 7546
Db 1202 ethrlealalystrhrtrpansglnprolyscysprothrmetileasptpillelyslu 1222
Oy 7547 AACATGGCAGATATACACCATGAAATATACATGACCCATTAAGAG-GATGAGTTCATC 7605
Db 1222 smetrhrpshlerythrmetyrlytrylalaalallelysnasnapglnpheillse 1242
Oy 7606 CTTTGACAGATATGATGAAAGCTGGAACCATCATCTCAGCAAACTAACACAAGAACA 7665
Db 1242 rPheValgllythrtrpmetlyslleuGlutThrlelleuenserlylseuerglnlgl 1262
Oy 7665 GAAACCAACACCACTGTCCTCACTGTAAAGTGGAGT 7705
Db 1262 nlystrlyshlsargllepheserleulleolyglyasn 1275

RESULT 3
ID 000360 PRELIMINARY: PRT: 1275 AA.
AC 000360.
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUN-1997 (TREMBlrel. 04, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative p150.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97285120; PubMed=9140393;
RA Sasaman D.M., Dombroski B.A., Moran J.V., Kimberland M.L., Naas T.P.,
RA Deberardinis R.V., Gabriel A., Swergold G.D., Kazazian H.H. Jr.;
RT "Many human L1 elements are capable of retrotransposition.";
RL Nat. Genet. 16:37-43(1997).
DR EMBL; U93563; AAC51261.1; -.
DR InterPro: IPR005135; Exo_endo_phos.
DR InterPro: IPR000477; RVTse.
DR Pfam: PF03372; Exo_endo_phos; 1.
DR Pfam: PF00078; Rvt; 1.
KW RNA-directed DNA polymerase.
SQ SEQUENCE 1275 AA; 149062 MW; 350B4F0358E525F0 CRC64;
```

```
US-10-083-853b-2 (1-29921) x 000360 (1-1275)
Oy 3428 ATGGCAGGATCAAAATTCACACATATAATATTAACTTAATGGTAAATGGCTAAATTC 3487
Db 1 MetthrlyserasenseniillethrilleuthrleuasnilleasnGlyleuasnser 20
Oy 3488 CCAATTAAGACACAGACTGCAAAATGGATTGAAGTCAAGACCCATGAGTGTCTGT 3547
Db 21 AlailelysarGhlsargargAlasertrpilleysserGlnaspproserValcyscys 40
Oy 3548 ATTCAGGAGGCCATCTCACATGAAAACACACATAGCTCCAAATTAAGGATGGAGG 3607
Db 41 lleGlnluthrHlsleuthrCysArgrasprhrHlsAargleuLysllelysllytrparg 60
Oy 3608 AAGATTTTCAAGTAAATGGAACCAAAAAAAGAGGAGGGTGAATCTGATCTCT 3667
Db 61 LyslletryGlnAlaasnglylsGln---LyslyslalaglyValaileleuValiser 79
Oy 3668 GATAAAGACACTTTAAACCAACAAGATCAAAAGACAAAGAGCCATTACATAATG 3727
Db 80 AsplystrhrasrphelyprothrlysllelysarGAsrlysglnGlyHlslyrilewet 99
Oy 3728 GTAAAGCATCAATGGAACAGAGAGCTACTATCCCTAAATPACATGACCCCAATACA 3787
Db 100 VallysglyserlleGlnlglndlndleuthrilleuasnilletrylalaProasnthr 119
Oy 3788 GGAGACCCAGATTCATTAAGCAAGATTCTTAAGACACTTAACAAAGACTTGACTCCAC 3847
Db 120 GlylarProalrphellelyslGlnValleuenserAsrpleuGlnArGAsrleuAspserHls 139
Oy 3848 ACAATATATGGGAGCTTAATAATTAATAGACACTTAAACCCCATCAATATTA 3907
Db 140 ThrleullemetGly-----AsrphesantthrProleuserthrleu 153
Oy 3908 GGCAATCAATAGACACAGAAATTAACAAGATATCCAGAGTGAATGACTGTGAC 3967
Db 154 AsparGserthrArGglnLyValaslnlysnasprhrGlnluleuasnserAlaleuHls 173
Oy 3968 CAAGCGAGCAATATAGATATATACAGAACTCCCAACCCCAATCAACAGATATACATC 4027
Db 174 GlnAlasrpleulleasprilletryrArgrhrleuHlsprolysserthGlnlythrph 193
Oy 4028 TTTCAGCATCATTAACCTATTTTAAATGACCTGTAATTTTAAGTAAACACTC 4087
Db 194 PheleuAlaProHlsHlsThrlyserlyslleasprHlsIleValglyserlyAlaleu 213
Oy 4088 CTCAGCAAAATGCAAAAGACAGAAATCTCTAACAAACAGTCTCAGACTACAGTGCATC 4147
Db 214 leuSerlyscyslysarGlythrGlnllethrHlsnlyrleuSerAsrHlsSerAlalle 233
Oy 4148 TATTTAGCACTCACAATTAAGAAACTCACTCAAAATCACAACATACATGGAACCTGAMC 4207
Db 234 LysleuGluleuArGlyllelyAsnleuthrlnseArGserthrThrlylyleuasn 253
Oy 4208 AACCTGCTCTGAATGACTACTGGTAAATTAACAATAAGCAAAAAATAAGATGTC 4267
Db 254 AsnleuLeuLeuAsnAsptrytrValHlsnslnglumeclysAlaGlnllyleuMerphe 273
Oy 4268 TTTGAACCAATGAGAACCAAGACAAATGACCAAGATCTCTGGGGCATTTTAAAGCA 4327
Db 274 PheGlnluthrAsnGlnAsnlysnAsprthrlyrGlnlnsnleuthrAsrAlaPheLyAla 293
Oy 4328 GTGTGTAGAGGAATTTATATGACTAGATGCTTACAAGAGAAAGCAGGAATATCTAAA 4387
Db 294 ValcysArGlylysrphelleleleuAsnAlaArlyrsArGlysglnGlnlArGserlyls 313
Oy 4388 ATAGACACCTTAACATCAACAATTAAGAAAGACTAGAGAAAGACAAACAAATTCAAA 4447
Db 314 llsasprhrleuthrSerGlnleuLyglndlndllysglnGlnlndlthHlsSerlyls 333
Oy 4448 GCTAGCAGAGACAGAAATTAACATCAAGATCAGACACAGAACCTGAAGGAGATGACACAA 4507
```

Dh 334 AlaSerArgArgIngluilethrLysIleArgAlaGluIleuLysGluIleGluThrGln 353
OY 4508 AAGCCCTCAATTAATCAATGAAATCCAGAGCTGGTTTGGAAAGATCGCAAAAT 4567
Dh 354 LysThrLeuIleuLys-IleasnGluSerArgSerTrpPheGluArgIleasnLysIl 373
OY 4568 ----AGACACTAGACACTAATTAAGAAAGAGAGAGAGAAATCAAGATGCAT 4623
Dh 373 eaSPArgProLeuAlaArgLeuIleLysLysArgGluLysAsnGlnIleasprHII 393
OY 4624 AAAAATGATAAAGGGATATCACACCGATCCACAGAAATCAAACTATATCAGA 4683
Dh 393 elYsAsnAspLysGlyAspIleThrAspProThrGluIleGlnThrIleArgG1 413
OY 4684 ATATTATAACCTCTATCAATTAATCAATAGAAATCTGAGAAATGATAATCTT 4743
Dh 413 uTrpTrpLysHisLeuTrpAlaAsnLysLeuGluAsnLeuGluLeuIleAspThrPheLe 433
OY 4744 GGACACATATGTAGCCCTGTATGAGACCTTGGGGACAGAACAAAAGGGGTGAATGCAGAA 4803
Dh 433 uAsp----- 434
OY 4804 ATAAAGACAAACAAACAAAGATATGTTTGGAAAGTAGGGGTCAAGGGCACTTGCTCT 4863
Dh 434 ----- 434
OY 4864 AATGACAAAGGGCCCTGAGCTTACACACCTCTGTATTTATTAGGCAAAAGAGATAGC 4923
Dh 434 ----- 434
OY 4924 GAGAGGTGAGTTGGAAGAGAGGTCACTGTAGTCCAGAGTAGCCCTGCAAGACTGC 4983
Dh 434 ----- 434
OY 4984 ATTCCCAACAATAGGCTTACATGTCCACAGTAGATAACCTCAAGAGCCAGTCCAGG 5043
Dh 434 ----- 434
OY 5044 GAGTGATGGCCCTCAGCAAACTTCTAGGGCAGGACAGAAAGTAAGTTGCCACATTC 5103
Dh 434 ----- 434
OY 5104 GATATCAGATAAACAAGTTTCTGTGTGATCAAGTACCTCAGTGGATGCTGAGTTGG 5163
Dh 434 ----- 434
OY 5164 TCATGATCCCTTGGCTTTTGGCTCCCAAAACACATACACCTCTCAAGACTAAACA 5223
Dh 435 -----ThrTyrThrLeuProArgLeuAsnG1 443
OY 5224 GGAAGAAGTCAATCCCTGATATACAGTAAAGTTCAAATTAATTAAGACAGTAATGA 5283
Dh 443 nGluGluValGluSerLeuAsnArgProIleThrGlySerGluIleValAlaIleIleas 463
OY 5284 TAGCCTACCAACAAAAAAGAGCCAGACGACAGAGGATTCACAGCCAAATTCACAGAG 5343
Dh 463 nSerLeuProThrLysLysSerProGlyProAspGlyPheThrAlaGluPheTyrGlnAr 483
OY 5344 GTCAAAAGAAAGCTGCTACTATTCCTTTCGAACATTCCTCAAAAATAA---AATGG 5400
Dh 483 gTyrLysGluGluLeuValProPheLeuLeuLysLeuPheGlnSerIleGluLysGlu-G 503
OY 5401 GAATCCCTCACTACTATTTAGAGAGCCAGCATCATCGATACCAAAACCTAGCAGTG 5460
Dh 503 LylleLeuProAsnSerPheTyrGluAlaSerIleIleLeuIleProLysProGlyArgA 523
OY 5461 ACAACAACAAAAGAGAAATTCAGGCCCATATCCCTGATGAACATTTGATGGAATTC 5520
Dh 523 sPThrThrLysLysGluAsnPheArgProIleSerLeuMetAsnIleAspAlaLysIleL 543
OY 5521 TCATATAAATACGTGGCAACCAATCCAGACAGACATCAACAAAAGCTTATCTACATATC 5580
Dh 543 euAsnLysIleLeuAlaAsnArgIleGlnIleHisIleLysLysLeuIleHisHisasp 563

OY 5581 AAGTGGCGTCATCCCTGGGATCGAAGCGCTGTTCAAAATATCAATCAATATATAG 5640
Dh 563 InValGlyPheIleProGlyMetGlnGlyTrpPheAsnIleValGlySerIleAsnValI 583
OY 5641 GCCATCACAATTAACAGAACCAATGACAAAACACATGATTATCTCAATGATGCAGAAA 5700
Dh 583 legInHisIleAsnArgAlaLysAspLysAsnHisMetIleIleSerIleAspAlaGluL 603
OY 5701 AGGCTTTGTCAAAATTCCAACAGCCCTTCATGCTTAAATTCACATTAACCTAGGTTCG 5760
Dh 603 ysaIaPheAspLysIleGlnGlnProPheMetLeuLysTrpIleAsnLysLeuGlyIleA 623
OY 5761 ATGGAATGATTCACAAAATATAAGACTTTATATC-AAACCCACAGCCCAATATCATAC 5819
Dh 623 spGlyThrTyrPheLysIleIleArgAlaIleTyrAspLysProThrAlaAsnIleArgL 643
OY 5820 TGAATGGGCAAAAACCTGGAAGCATTCCTTGGAACTGGCACAGACAAAGATGCCCTC 5879
Dh 643 euAsnGlyGlnLysLeuGluAlaPheProLeuLysTrpGlyThrArgGlnGlyCysProL 663
OY 5880 TCTCACACCTCTATTCAATATCTATGGAAGTTCTGGCCAGGCGCATAGCAATAGA 5939
Dh 663 euSerProLeuLeuPheAsnIleValLeuGluValLeuAlaArgAlaIleArgGlnGluL 683
OY 5940 AAGAAATTAAGGGATTCAAATAGAAAGAGAGAGAGATCATATTGCTCTGTTCAGATG 5999
Dh 683 yseGluIleLysGlyIleGlnLeuGlyLysGluGluValLysLeuSerLeuPheAlaAspA 703
OY 6000 ACATGTTGTATATTAGAAACCCATCGCTCAGCCCAAAAACCTCCTTAAGCTGATAA 6059
Dh 703 spMetIleValTyrLeuGluGluAsnProIleValSerAlaGlnAsnLeuLeuLysLeuL 723
OY 6060 GCAACTTCACAAAGTCTCAGACACACAAATCAATGTGCAAAAATCAAGACTTCTAT 6119
Dh 723 eraSnPheSerLysValSerGlyTyrLysIleAsnValGlnLysSerGlnAlaPheLeuTr 743
OY 6120 ACCGCAATTAAGACAAACAGAGAGAGCCAAATCATGAGTGAAGCTCTACAAATTGCTA 6179
Dh 743 yTrpHisAsnAlaArgInThrGluSerGlnIleMetSerGluLeuProPheThrIleAlas 763
OY 6180 CAAGAGAAATTAATACCTTAGAATACACTTCAAGAGACAGCTAGAACTTCAAGG 6239
Dh 763 erLysArgIleLysTyrLeuGlyIleGlnLeuThrArgAspValLysAspLeuPheLysG 783
OY 6240 AGAACTTCACAACATGATCAAGAAATAGAGAGACACAAACCAATATGAAAAACTTC 6299
Dh 783 LuAsnTyrLysProLeuLeuLysGluIleLysGluAspTrpHisLysTrpLysAsnLeP 803
OY 6300 CATGCTCAGATAGTAAAGATCAT-----GAAATGGCATATCCCAAGTAATTT 6352
Dh 803 rocYsSerTrpValGlyArgIleAsnIleValLysMetAlaIleLeuProLysValIleTr 823
OY 6353 ATAGATTCAAGTGTACCCCATCAAGCTACATGACTTTCTTACAGAAATGAAAAA 6412
Dh 823 yTrArgPheAsnAlaIleProIleLysLeuProMetThrPhePheThrGluLeuGluLysTr 843
OY 6413 CAACTTAATTCATATGAGACCAAAAAGAGCCACAGACGCAAGACAAATCTTAAGC 6472
Dh 843 hrThrLeuLysPheIleTrpAsnGln-LysArgAlaLysIleAlaLysSerIleLeuSer 862
OY 6473 AAAAAGAAACACTGGAGGTATCATGCTACCTGACTTAACATATCTAAGGCTACA 6532
Dh 863 GlnLysAsnLysAlaGlyIleThrLeuProAspPheLysLeuTyrTrpLysAlaThr 882
OY 6533 GTAACCAAACTGCATGCTGCTGACCAAAACAGATATATAGCAATGGAACAGACA 6592
Dh 883 ValThrLysThrIleArgTrpTyrTrpGluAsnArgspIleAspGlnTrpAsnArgTrH 902
OY 6593 GAGACCTCAAGAAATTAACAT-GCAATCTACATCCATCTGTCTTGAACAAACCTGACAAA 6651
Dh 903 GluProSerGluIleThrProHisIleTyrAsnTyrLeuIlePheAspLysProGluLys 922

QY 3908 GGCAGATCAATGACAGAAATTTAAACAGGATATCCAGAGTTGAAGCTGCTGCAC 3967
||||| ||||||| ||||||| ||||||| |||||||
Db 154 AsparGSerThrArgGlnLysValAsnLysAspThrGlnGluLeuAsnSerAlaLeuHis 173
QY 3968 CAAGCGACCTAATAGATATCTACAGAACTCCACCACCCCAATCAACAGATATACATC 4027
||||| ||||||| ||||||| ||||||| |||||||
Db 174 GlnAlaAspLeuIleAspIleTyrArgThrLeuHisProLysSerThrGluTyrThrPhe 193
QY 4028 TTCTGACATCAATACACCTATTTTAAATGACCATGTATTTTAACTAAACATC 4087
||||| ||||||| ||||||| ||||||| |||||||
Db 194 PheSerAlaProHisThrTyrSerLysIleAspHisIleValGlySerLysAlaLeu 213
QY 4088 CTCAGCAATGCAAAAGACAGAAATTCCTAAACAGCTCTCAGACCTACATGCAATC 4147
||||| ||||||| ||||||| ||||||| |||||||
Db 214 LeuSerLysCysLysArgThrGluIleIleThrAsnTyrLeuSerAspHisSerAlaIle 233
QY 4148 TATTTGACATCAATTAAGAACTCCTCAAAATCAACAATCATCATGGAATGTGAC 4207
||||| ||||||| ||||||| ||||||| |||||||
Db 234 LysIleGluLeuArgIleLysAsnLeuThrGlnSerArgSerThrThrTyrLysLeuAsn 253
QY 4208 AACCTGCTCTGATGACTACTGGTAAATTAACAAATGAGCAAAATTAAGATGTC 4267
||||| ||||||| ||||||| ||||||| |||||||
Db 254 AsnLeuLeuLeuAsnAspTyrTyrValHisAsnGluLeuLysAlaGluIleLysMetPhe 273
QY 4268 TTTGAAACCAATGAGAAACAAGACACATGTACAGATCTCTGGGCATATTTAAAGCA 4327
||||| ||||||| ||||||| ||||||| |||||||
Db 274 PheGlnThrAsnGluAsnLysAspThrThrTyrGlnAsnLeuThrAspAlaPheLysAla 293
QY 4328 GTGTGTAGAGGAAATTTATAGCATAGATGCTTACAGAGAAAGCAAGAAATATCTAAA 4387
||||| ||||||| ||||||| ||||||| |||||||
Db 294 ValCysArgGlyLysPheIleAlaLeuAsnAlaTyrLysArgLysGlnGluArgSerLys 313
QY 4388 ATGACACCTTACATCAATTAAGAACTAGAGAAAGCAAGCAAAATTCAAA 4447
||||| ||||||| ||||||| ||||||| |||||||
Db 314 IleAspThrLeuThrSerGlnLeuLysGluLeuGlnLysGlnGlnTyrHisSerLys 333
QY 4448 GCTAGCAGAGACAAAGAAATTAAGATCAGAGCAAGCAACTGAAGATAGAGACAA 4507
||||| ||||||| ||||||| ||||||| |||||||
Db 334 AlaSerArgGlnGlnGluIleThrLysIleArgAlaGluLeuLysGluIleGlnThrGln 353
QY 4508 AAAGCCCTTCAATTAATCAATGATCCAGGACCTGTTTTGAAAGATCAGCAAAAT 4567
||||| ||||||| ||||||| ||||||| |||||||
Db 354 LysThrLeuGlnLysIleAsnGlnSerArgSerThrPheGlnLysArgLysIleAsnLys 373
QY 4568 ----AGACCACTAGACAGACTAATTAAGAAAGAGAGAGAAATCAAGATGCAAT 4623
||||| ||||||| ||||||| ||||||| |||||||
Db 373 AspArgProLeuAlaArgLeuIleLysLysLysArgGluLysAsnGlnIleAspThrIle 393
QY 4624 AAAAATGATAAAGGGATATCACACCGATCCACAGAAATCAAACTTTATTCAGAGA 4683
||||| ||||||| ||||||| ||||||| |||||||
Db 393 GluAsnAspLysGlyAspIleThrThrAspProThrGlnIleGlnThrThrIleArgGlu 413
QY 4684 ATTTTATTAACACCTCTATGCAATTAATACTAGAAATCTAGAAAGAAATGATTAATTCCT 4743
||||| ||||||| ||||||| ||||||| |||||||
Db 413 utyTyrLysHisLeuTyrAlaAsnLysLeuGlnAsnLeuGlnIleMetAspThrPheLe 433
QY 4744 GGACACATATGTAGCCGTATGAGACCTTGGGGGACAGAACAAAAGGGGTGATGCGAA 4803
||||| ||||||| ||||||| ||||||| |||||||
Db 433 uAsp----- 434
QY 4804 ATAAAGACAAAGACAAAGATATGTTTGAAGTAGGGGTCAAGGGGCAATCTGCCTCT 4863
||||| ||||||| ||||||| ||||||| |||||||
Db 434 ----- 434
QY 4864 AATGCAAGAGGCCCTGAGCTTACACACCTCTGTATTTATTAGCAAAAGAGATAGC 4923
||||| ||||||| ||||||| ||||||| |||||||
Db 434 ----- 434
QY 4924 GACAGGGTGAGTTGGAAGAGAGGTACAGCTGTAGTCCAGAGTAGGCTGCAAGACTGC 4983
||||| ||||||| ||||||| ||||||| |||||||
Db 434 ----- 434

QY 4984 ATTCTCAAAACAATAGGCTCTAGATGTCCAGTAGAATAACCTGAAGAGCCAGTGCAGG 5043
----- 434
Db 434 ----- 434
QY 5044 GAGTAGAGCCCTCAGCAAACTTTAGGCGAGGCACAGAAAGTAAGTTGCCACATCTCT 5103
----- 434
Db 434 ----- 434
QY 5104 GTATTCACGATTAACAGTTTGCTGTGTGATCAAGTACCCCTCAGTGAATGCTGAGTTGG 5163
----- 434
Db 434 ----- 434
QY 5164 TCATGATCCCTTTGGCCTTTTGGCCCAAAACATACACCTCTCAAGATCAAAACA 5223
----- 435
Db 435 ----- 435
QY 5224 GGAAGAAGTCAATCCCTGATATATACCAGTACAAAGTTCAAAATGAGCAAGTAATGGA 5283
||||| ||||||| ||||||| ||||||| |||||||
Db 443 ngIuGluValGluSerLeuAsnArgProIleThrGlnGlySerGluIleValAlaIleIleAs 463
QY 5284 TAGCTTACCAACCAAAAAAGTCCAGAGACAGACCGATTTGACAGCCCAATTCACAGAG 5343
||||| ||||||| ||||||| ||||||| |||||||
Db 463 nSerLeuProThrLysSerProGlyProAspGlyPheThrAlaGluIlePheTyrGlnAla 483
QY 5344 GTACAAAGAGAGCTGCTACTATCTCTTGAACCTTTTCCAAAAATAGAA--AATGG 5400
||||| ||||||| ||||||| ||||||| |||||||
Db 483 gTyrLysGluGluLeuValProPheLeuLeuLysLeuPheGlnSerIleGluLysGlu 503
QY 5401 GAATCTCTCACTATTTTACAGGCGACATCATCTGATACCAAACTAGACAGTG 5460
||||| ||||||| ||||||| ||||||| |||||||
Db 503 LysIleLeuProAsnSerPheTyrGluValaSerIleIleLeuIleProLysProGlyArg 523
QY 5461 ACACAAACAAAGAGAAATTTCAAGGCCATATCCCGATGCAATGATGATGAAATCC 5520
||||| ||||||| ||||||| ||||||| |||||||
Db 523 spThrThrLysLysGluAsnPheArgProIleSerLeuMetAsnIleAspAlaLysIle 543
QY 5521 TCAATAAATACGTGGCAAAACCAATCCAGAGACATCAAAAGCTTATACCATGATC 5580
||||| ||||||| ||||||| ||||||| |||||||
Db 543 euAsnLysIleLeuAlaAsnArgIleGlnGlnHisIleLysLysLeuIleHisIleAspG 563
QY 5581 AAGTGGCGTCATCCCTGGGATGCGAAGCGTGGTTCAAAATATGCAATCAATTAATGTAG 5640
||||| ||||||| ||||||| ||||||| |||||||
Db 563 lValGlyPheIleProGlyMetGlnGlyTrpPheAsnIleArgLysSerIleAsnValI 583
QY 5641 GCCATCACAATAACAGAACCAATGACAAAAACCAATGATATCTCAATGATGACAGAAA 5700
||||| ||||||| ||||||| ||||||| |||||||
Db 583 leGlnHisIleAsnArgAlaLysAspLysAsnHisMetIleIleSerIleAspAlaIle 603
QY 5701 AGGCTTTGTCAAAATTCACACAGCCCTTCATGCTAAATTCACATTAATGATGCTATCG 5760
||||| ||||||| ||||||| ||||||| |||||||
Db 603 ysaIaPheAspLysIleGlnGlnProPheMetLeuLysThrIleAsnLysLeuGluIleAla 623
QY 5761 ATGGAATGTATCTCAAAATTAAGAGCTATTATAC-AAACCCACAGCCCAATATCATAC 5819
||||| ||||||| ||||||| ||||||| |||||||
Db 623 spGlyThrTyrPheLysIleIleArgAlaIleTyrAspLysProThrAlaAsnIleIle 643
QY 5820 TGAATGGCAAAACCTGAGAGCATTCCTCTTGAAGAACTGGCACAAGCAAGATGCCCTC 5879
||||| ||||||| ||||||| ||||||| |||||||
Db 643 euAsnGlnGlnLysLeuGlnAlaPheProLeuLysThrGlyThrArgGlnGlyCysPro 663
QY 5880 TCTCACACATCTATTAAGATACTATTTGAAGTCTGGCGAGGCAATGACAGCAATPAGA 5939
||||| ||||||| ||||||| ||||||| |||||||
Db 663 euSerProLeuLeuPheAsnIleValIleuGluValIleuAlaAlaGlnAlaIleArgGlnIle 683
QY 5940 AAGAAATAAAGGATATCAATAGAAAGAGAGAGCAAGTCAATATGCTCTGTTGCAGATG 5999
||||| ||||||| ||||||| ||||||| |||||||
Db 683 ysgIuIleLysGlyIleGlnLeuGluLysGlnIuValLysLeuSerLeuPheAlaAsp 703
QY 6000 ACATGTTGTATATTATTAAGAAACCCCATGCTCAGAGCCAAAAATCTTAACCTGATTA 6059
||||| ||||||| ||||||| ||||||| |||||||
Db 703 spMetIleValTyrLeuGlnAsnProIleIleSerAlaGlnAsnLeuLysLeuIle 723
QY 6060 GCAACTTCAGCAAGTCTCAGACACCAAAATCAATGTGCAAAATCAAGCAAGTCTTAT 6119

Db 723 eTasnheserlyValSerGlyTyrLysIleasnValGlnLysSerGlnAlaPheLeuT 743
 QY 6120 AGCCCAATATAGACAACAAGAGAGCAATCATGAGTGGATGCTCATGATGCTGATGCTA 6179
 Db 743 yTThrasnsnargIntnrgIuserGlnIleMetLylGlnLeuProPheThrIleAlaS 763
 QY 6180 CAAGAAGATTAATACCTAGGAATACAATTACAAGGACACAGTGAAGACTTCTCAAG 6239
 Db 763 eTlysrArgIleLysTyrLeuGlyIleGlnLeuThrArgAspValLysAspLeuPheLysG 783
 QY 6240 AAGACTNCAACCACTGATCAAGGAATAAGAGAGACACAAACAATGGAATAATC 6299
 Db 783 lAsnTyrLysProLeuLeuLysGlnIleLysGlnAspThrAsnLysTyrLysAsnIleP 803
 QY 6300 CATGCTCACAGATAGTAAGAATCAT-----GAATAGCCATACCTAGCCCAAGTAAAT 6352
 Db 803 rOCysSerTrrpValGlyArgIleAsnIleValLysMetAlaIleLeuProLysValIleT 823
 QY 6353 ATAGATTAGTGTACCCCAATCAAGCTACCATGACTTCTTACAGAAATTGGAATAA 6412
 Db 823 yTArgPheAsnAlaIleProIleLysLeuProMetThrPhePheThrGlnLeuGlnLysT 843
 QY 6413 CAACCTTAATTCATATGGAACCAAAAAGGCCACAGAGCCCAAGCAATCTTAAGC 6472
 Db 843 hTThrLeuLysPheIleTrrpAsnGln-LysArgAlaArgIleAlaLysSerIleLeuSer 862
 QY 6473 AAAAGAACAAAGCTGGAGTATCATGCTGAGTAAACCTATATCATATAGAGCTACA 6532
 Db 863 GlnLysAsnLysAlaGlyGlyIleThrLeuProAspPheLysLeuTyrLysAlaThr 882
 QY 6533 GTAACCAAAATGTCATGCTACTGTGTACAAACAGATATATAGACCAATGGAACAGACA 6592
 Db 883 ValThrLysThrAlaTrrpTrrpTrrpGlnAsnArgAspIleAspLIntnrgAsnArgThr 902
 QY 6593 GAGACCTGGAATTAAC-ACAGCAATCTCATCCATCCATGATCTTACCAAACTGCAAA 6651
 Db 903 GlnProSerGlnIlePheProHisIleTyrAsnTyrLeuIlePheAsnLysProGlnLys 922
 QY 6652 AACAGCAATGGAAGAAAGATTCCTATTAATTAATGCTGTGGAAGAAAGCTGGATGCC 6711
 Db 923 AsnLysGlnTrrpGlyLysAspSerLeuPheAsnLysTrrpCysTrrpLusnTrpLeuAla 942
 QY 6712 ATATGCAAGAACTGAACTGATCCCTTCTTACACCTTATACAAAGTAACTACAGACA 6771
 Db 943 IleCysArgLysLeuLysLeuAspProPheLeuThrProTyrThrLysIleAsnSerArg 962
 QY 6772 TGAATTAAGACTTAATATAGACATAAACCATTAACCAACCA-GAAGAAAACCTAGGC 6830
 Db 963 TrpIleLysAspLeuAsnValArgProLysTrnIleLysThrLeuGlnLusnLeuGly 982
 QY 6831 AATACCATTCAGATATAGACATGGCAAGCAATTCATGACTTAACCAACCAAGCAATG 6890
 Db 983 IleThrIleGlnAspIleGlyValGlyLysAspPheMetSerLysThrProLysAlaMet 1002
 QY 6891 GCAACAAAAGCCAAAATAGACAAGTGGGATGCTGATTAACCTATAGAGCTTGGACAGACA 6950
 Db 1003 AlatrLysThrLysIleAspLysTrrpAspLeuIleLysLeuLysSerPheCysTrnAla 1022
 QY 6951 AAAAAAATCTGATCAGAGTGAACAGCAACCTACAGAAATGGAGAAAATTTTTCAT 7010
 Db 1023 -LysGlnThrThrIleAsnGlnValAsnArgGlnProThrThrTrpGlnLysIlePheAlaTh 1042
 QY 7011 CTATGATCTGCAAAAGCTAATATCCAGATCTACGAAGAATTAACCAATTTTACAA 7070
 Db 1042 TrrpSerSerAspLysGlyLeuIleSerArgIleTyrAsnGlnLysLeuGlnIleTyrLys 1062
 QY 7071 GAAAAA-----AACAAACCCGTCAAAATATAGGCAAGATATAGACAGACACTTGCAAA 7126
 Db 1062 slsLysThrAsnAsnProIleLysLysTrrpAlaLysAspMetAsnArgHisPheSerLys 1082
 QY 7127 AGAAGACATTTATGAGCCAAACAAACATATGAAAAAACCCTATCATCTTGTGCTGTAG 7186
 |||||||

Db 1082 sGlnAspIleTrrpAlaAlaLysLysHisMetLysLysCysSerSerLeuAlaIleAr 1102
 QY 7187 AGAATGCAAAACCAACCAAGTGCATACCATCTCAAGTATGATGATGATGATGATC 7246
 Db 1102 gGlnMetGlnIleLysThrThrMetArgTrrpHisLeuThrProValArgMetAlaIleI 1122
 QY 7247 TAAAAATCAGCAACCAACCAATGCTGAGAGAGATGTGGAATAATGGAACACTTTTCCA 7306
 Db 1122 eLysLysSerGlyAsnAsnArgCysTrrpArgGlyCysGlyGlnIleGlyThrLeuLeuH 1142
 QY 7307 CTGTGCTGGGAGATGAATTAATGTTAGTCAACCATGTGGAAGACAGCTGGAGATTCCTTAA 7366
 Db 1142 scYsTrrpTrrpAspCysLysLeuValGlnProLeuTrrpLysSerValTrrpPheLeuAr 1162
 QY 7367 GGAATGCAACCAAGATTAATATCATTTGACCCAGCAATCCCATTAAGTATATACCAAA 7426
 Db 1162 gAspLeuGlnLeuGlnIleProPheAspProAlaIleProLeuLeuGlyIleTrrpLys 1182
 QY 7427 GGAATTAATATCATTTTATTAAGACACATGACACATATGTTTATGACGACTGAT 7486
 Db 1182 sAspTrrpLysSerCysCysTrrpLysAspThrCysTrrpArgMetPheIleAlaLeuPh 1202
 QY 7487 CACAATGCAAGAGCTTGAACCAACCAAGTCCATGATGATGATGATGATGATGATGATGAT 7546
 Db 1202 eThrIleAlaLysThrTrrpAsnGlnProLysCysProThrMetIleValTrrpLysLys 1222
 QY 7547 AACATGCAATATACCATGCAATATCATGATGATGATGATGATGATGATGATGATGATGAT 7605
 Db 1222 smetTrrpHisIleTrrpThrMetCylTrrpAlaAlaIleLysAsnAspGlnPheIleSe 1242
 QY 7606 CTTTGCAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7665
 Db 1242 rPheValGlyThrTrrpMetLysLeuGlnThrIleIleLeuSerLysLeuSerGlnGln 1262
 QY 7666 GAAACCAACCAACCAACATGTTCTCACTTGAATGGGAGCT 7705
 Db 1262 nLysThrLysHisCysIlePheSerLeuIleGlyLysn 1275
 |||||||

RESULT 5
 Q9UN80 PRELIMINARY; PRT: 1275 AA.
 AC Q9UN80:
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical 149.0 kDa protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-RETROTRANSPOSON L1;
 RX MEDLINE=98361157; PubMed=9697692;
 RA Schwahn U., Lenzner S., Dong J., Fell S., Hinzmann B.,
 RA van Duinhoven G., Kirschner R., Hemberger M., Bergen A.A.,
 RA Rosenberg T., Pinkers A.J., Fundele R., Rosenthal A., Cremers F.P.,
 RA Ropers H.H., Berger W.;
 RT "Positional cloning of the gene for X-linked retinitis pigmentosa 2.";
 RT Nat. Genet. 19:327-332(1998).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-RETROTRANSPOSON L1;
 RA Kazazian H.H.;
 RT "Full-length human L1 insertions retain the capacity for high
 RT frequency retrotransposition in cultured cells.";
 RT Submitted (May-1999) to the EMBL/GenBank/DBD databases.
 DR EMBL: AF148856; AAD39215.1; -
 DR InterPro: IPR005135; Exo_endo_phos.
 DR InterPro: IPR000477; RVTse.
 DR Pfam: PF03372; Exo_endo_phos; 1.
 DR pfam: PF00078; rvt; 1.


```
Db 503 LylleupProaenSerPheTyrgLualasellelleuileProlysProGlyAArg 533
QY 5461 ACACAACAAAAGAGAAATTTTCAGGCCATTCCTGATGAACATGATGTAATCC 5520
Db 523 spthrThrylsySgluasnPhearProIleSerleuMetasnIleasPAllysIleL 543
QY 5521 TCATAAATAATCTGGCAAAACCAATCCAGAGACATCAAAAAGTTATCTACATGATC 5580
Db 543 euasnlySleleuAlaasnArgIleGlnGlnHisIlelySlysleuIleHisIsapG 563
QY 5581 AACTTGGCGTCATCCTGGATGTCAGGCTTGCTCAAAATATCTCAATATCAATAGTAG 5640
Db 563 lvalSlyPheIleProGlyMetGlnGlyTrpPheasnIleArgGlySerIleasnValI 583
QY 5641 GCCATCAATAAACAGAACCAATGACAAAACACATGATTCATCTCAATGATGACAAA 5700
Db 583 lGlnHisIleasnAlaArgAlaAspLysasnHisMetIleIleSerIleasPAlleGUL 603
QY 5701 AGGCTTTGTCAAAATTCACAGCCCTTCATGCTAAAAATTCAGTAAACTAGTAGTCG 5760
Db 603 ysAlaPheaspLysIleGlnGlnProPheMetleuysThrIleuasnLysleuGlyIleA 623
QY 5761 ATGGATGTATCTCAAAATTAATAGACTATTATAC-AAACCCACAGCAATATCATAC 5819
Db 623 spElyThrTyPheIleIleArgAlaIleTyraSpLysProThrAlaasnIleIleL 643
QY 5820 TGAATGGCAAAACATGAAAGCTTCCTTTGAGATGCGACACAAAGATGCGCTC 5879
Db 643 euasnLysGlnLysleuGlnAlaPheProleuysThrGlyThrArgGlnGlyCysProl 663
QY 5880 TCTCACCACTCCTATTCAAGATACTATTGGAATCTGCGCCAGGCAATCAGGCAATAGA 5939
Db 663 euserProleuLeuPheasnIleValleuGlnValleuAlaArgAlaIleArgGlnGluL 683
QY 5940 AAGAATTAAGGCTATCAATAGAAAGAGGAATCATATGCTGCTGTTGCGATG 5999
Db 683 yseGlnLysGlyIleGlnleuGlyLysGlnValLysleuSerleuPheAlaSpA 703
QY 6000 ACATGTTGTATATTAGAAAACCCATCGCTCAGGCCAAAACCTCTTAACCTGATTA 6059
Db 703 spMetIleValTyrlleuGlnuasnProIleValSerIleGlnasnleuLysleuIleS 723
QY 6060 GCACCTTCAGCAAAAGCTCAGGACACAAAATCAATGTGCAAAAATCACAAGCTTTAT 6119
Db 723 etasnPheSerIysValSerGlyTyrlLysIleasnValGlnLysSerGlnAlaPheLeuT 743
QY 6120 ACCCATATATAGACAAACAGAGAGCCAAATCATGAGTGAACCTCATTCACATGCTA 6179
Db 743 yThrTrasnasnArgInThrGlnSerGlnIleMetGlyLuleuProPheThrIleAlas 763
QY 6180 CAAAGAGAAATAAATACCTAGGAATACAACTTACAAAGGACAGACTGGAAGCTTCAAG 6239
Db 763 etLysArgIleLysTyrlleuGlyIleGlnleuThrArgaspValLysaspLeuPheLysG 783
QY 6240 AGACACTCAACACATCATCAGAGAAATAAGAGAGACACAAACAATGTGAAAAACATTC 6299
Db 783 luasnTyrlLysProleuLysGlnIleLysGlnIleThrasnLysTrpLysasnIleP 803
QY 6300 CATGCTCACAGATAGTAAGAAATCAT-----GAAATGGCATACAGCCCAAGTAAT 6352
Db 803 rocYserTrpValGlyArgIleasnIleValLysMetAlaIleleuProLysValIleT 823
QY 6353 ATAGATTAGTGTGATCCCATCATCAAGCTTCACTTGAATTCCTTCACAGAAATGGAAGAA 6412
Db 823 yThrArgPheasnAlaIleProIleLysleuProMetThrPhePheThrGlnleuGlnLysT 843
QY 6413 CAACTTTAAATTCATATGGAACCAAAAAGAGCCACAGAGCCAAAGCAATCTTAAGC 6472
Db 843 hThrIleuLysPheIleTrpAsnGln-LysArgAlaArgIleAlaLysSerIleleuSer 862
QY 6473 AAAAGAACAAAGCTGAGTATCATGACCTGAACTTAAACTATACATTAAGCTTACA 6532
Db 863 GlnLysasnLysAlaGlyIleThrLeuProaspPheLysleuTyrlLysAlaThr 882
QY 6533 GTAACCAAAACCTGATGTAAGTGTACCAAAAACAGATATATAGACCAATGAGACAGACA 6592
Db 883 ValThrLysThrAlaTrpTyrlTrpArgInasnArgAspLysPAlnThrPasnArgThr 902
QY 6593 GAGACCTCAGAAATTAAC-AC TGCAATCTACATCCATCTGATCTTTGACAAACCTGACAAA 6651
Db 903 GluProSerGlnIleMetProHisIleTyraSnTyrlleuIlePheaspLysProGlnLys 922
QY 6652 AACACAGCATGGAAGAAAGATTCCTATTTAAATAGGATGGTGAAGAAACCTGAGGC 6711
Db 923 AsnLysGlnTrpLysaspSerleuPheasnLysTrpCysTrpLusnTrpIleuAla 942
QY 6712 ATATGCAAGAAAGCTGAAATGATGATCCCTTCCTTACACCTTATACAAAGTAACTCAGACA 6771
Db 943 lIecysArgLysleuLysleuasnPropheleuThrProTyrlThrylsIleasnSerArg 962
QY 6772 TGAATTAAGACTTAATATTAAGACATTAACCATTAACCA-CAAGAAACCTAGGC 6830
Db 963 TrpIleLysaspLeuasnValLysProLysThrIleLysThrleuGlnLusnleuGly 982
QY 6831 AATACCATTCAGGATATGGACATAGGCAAAAGCTTCATGACTTAAACACCAAGCAAG 6890
Db 983 lIeThrIleGlnaspIleGlyValGlyLysaspPheMetSerLysThrProLysAlaMet 1002
QY 6891 GCACAACAAAGCCAAATATAGACAAGTGGATCTGATTAACTATAGACTTTCACAGACA 6950
Db 1003 AlaThrLysaspLysIleasnLysTrpAspLeuIleLysleuLysSerPheCysThrAla 1022
QY 6951 AAAAAAACTGTCATCAGAGTGCACAAGCAACCTACAGAAATGGAGAAAATTTTGCAT 7010
Db 1023 -LysGlnThrThrIleArgValaasnArgInProThrTrpGlnLysIlePheAlaTh 1042
QY 7011 CTATGATGTGCAAAAGCTTAATATATCCAGATCTCAGAAAGCTTAAACAAATTAACA 7070
Db 1042 rTyrlSerSerAspLysGlyLysleuIleSerArgIleTyraGlnleuLysGlnIleTyrl 1062
QY 7071 GAAAAA-----AACACCCCGTCAAAATATGGCAAGATATGACGACGACACTTCAAA 7126
Db 1062 sLysLysTrnAsnAsnProIleLysLysTrpAlaLysAspMetAlaArgHisPheSerly 1082
QY 7127 AGAAGACATTTATGCGCAACAAACATATGAAAAAACCCTATCATTCATGCTGTAG 7186
Db 1082 sGlnAspIleTyrlAlaLysLysHisMetLysLysCysSerSerleuAlaIleAr 1102
QY 7187 AGAAATGCAAAACAAACACAGTGCATACCATCATCAGTATGATGATGATGATGATC 7246
Db 1102 gGlnMetGlnIleLysThrThrMetArgTyrlHisleuThrProValArgMetAlaIleTl 1122
QY 7247 TAAAAAGTCAGAAACACAAATGTGAGAGAGATGAGAAATAGAACACTTTTCCA 7306
Db 1122 etLysLysSerGlyAsnAsnArgCysTrpArgGlyCysGlyGlnIleGlyThrleuLeuH 1142
QY 7307 CTGTGTGGTGAATGTAATTAATGATTCACATTTGAGACAGATGAGATGAGATTCCTTA 7366
Db 1142 scYsTrpTrpAspCysLysleuValGlnProleuTrpLysSerValTyrlArgPheLeuAr 1162
QY 7367 GATCTAGAACCAAGAAATTCATTTGACCCAGCAATCCCATTAAGTATATACCAAA 7426
Db 1162 gasPleuGlnleuGlnIleProPheaspProAlaIleProleuLysGlyIleTyrlProAs 1182
QY 7427 GGAATTAATATCTATTAATTAAGACACATGCACACATGATGTTATGACGACTGAT 7486
Db 1182 ngLutTyrlLysSerCysTyrlLysaspPhrCysThrArgMetPheIleAlaAlaLeuPh 1202
QY 7487 CACATAGCAAGAGCTTGACACCAACCAAAATGTCATAGTATGATGATGATTAAGAA 7546
Db 1202 eThrIleAlaLysTrnTrpAsnGlnProLysCysProThrMetIleasPTrpIleLysLy 1222
QY 7547 AACATGCGACATATCACCATCAATATCATGACAGCAATAAAG-GATGAGCTTCAATGC 7605
Db 1222 sMetTrpHisIleTyrlThrMetGlnTyrlAlaAlaIleLysasnaspGluPheIleSe 1242
```


QY 4984 ATTCTCAACAATAGGCTCTAGATGTCCAGTAGATTAACCTCAAGAGCCAGTGCAGG 5043
Db 434 ----- 434
QY 5044 GAGTGAATGGCCCTCAGCAAAACCTTTAGAGGACGACAGAAAGTAAGTTGGCCACATTC 5103
Db 434 ----- 434
QY 5104 GTATTCAGTAACACAGTTTGTCTTGTGATCAAGTAGAGCTCAGTGGAAATGCTGAGTTGG 5163
Db 434 ----- 434
QY 5164 TCATGATCCCTTTGGCTTTTGGCTCCCAAAACATACACCCCTTCAAGATAACCA 5223
Db 435 -----ThyTyrThrLeuProArgLeuasnG 443
QY 5224 GGAAGAAGTCAATCCCTGAATATACAGTAAGTAAGTCAATTAATGAAGCAATATATGA 5283
Db 443 nGluGluValGluSerLeuasnArgProIleThrGlySerGluIleValAlaIleIleAs 463
QY 5284 TAGCTTACCAACCAAAAAAGTCCAGAGCAGACGAGATTCACAGCCAAATTCACAGAG 5343
Db 463 nSerLeuProThrLysLysSerProGlyProAspGlyPheThrAlaGluPheTyrGlnArg 483
QY 5344 GTACAAAGAGAGAGCTGGTACTATTCCTTGAAGCTATTCCAAAAAATAGAA--AATGG 5400
Db 483 gTyrLysGluGluLeuValProPheLeuLeuLysLeuPheGlnSerIleGluLysGlu-G 503
QY 5401 GATCCCTCCCTAATCATTTTACGAGGCCAGCATCATCTGATACCAAAACCTAGACAGT 5460
Db 503 LysLeuProAsnSerPheTyrGluAlaSerIleIleLeuIleProLysProGlyArg 523
QY 5461 ACACAACAAAAGAGGAATTTTCAGGCCCATATCCCTGATGAACATTGATGTGAATGCC 5520
Db 523 spThrThrLysLysGluAsnPheArgProIleSerLeuMetAsnIleAspAlaLysIleIle 543
QY 5521 TCATTAATAATCTGGCAAAACCAATCCAGACGACATCAAAAAGCTTATCTACCATGATC 5580
Db 543 euAsnLysIleLeuAlaAsnArgIleGlnGlnHisIleLysLysLeuIleHisIleAspG 563
QY 5581 AAGTTGGCGTATCCCTGGGATGACAGGCGTGCCTCAAAATATCAAAATCAATTAAGTGA 5640
Db 563 IValAlaLysPheIleProGlyMetGlnGlyTyrPheAsnIleArgLysSerIleAsnValI 583
QY 5641 GCCATCACATAACAGAACCAATGACAAAACACATGATTTATCTCATAGATGACAGAA 5700
Db 583 IeGlnHisIleAsnArgAlaLysAspLysAsnHisMetIleIleSerIleAspAlaGlu 603
QY 5701 AGGCTTTGTCAAAATTCACAGCCCTTCATGCTAAATAATCTCAGTAACATAGTATCG 5760
Db 603 ysaIaPheAspLysIleGlnGlnProPheMetLeuLysThrLeuAsnLysLeuGlyIleA 623
QY 5761 ATGGATGTATCTCAAAATATTAAGAGCTATTATATC-AAACCCACAGCCATATATATC 5819
Db 623 spLysThrTyrPheLysIleIleArgAlaIleTyrAspLysProThrAlaAsnIleIle 643
QY 5820 TGAATGGCAAAAACGAGAGCATTCCTTGGAGACTGGCACAGACAGATGCGCTC 5879
Db 643 euAsnGlyLysIleGlnLeuGluAlaPheProLeuLysThrGlyThrArgGlnGlyCysPro 663
QY 5880 TCTCACCACTCTATTCAGATTAATCTGGAAGTTCGGCCAGGGCAATCAGGCATATGA 5939
Db 663 euSerProLeuLeuPheAsnIleValLeuGluValLeuAlaArgAlaIleArgGlnGlu 683
QY 5940 AAGAAATAAAGGTATATCAATAAGAAAGAGAGAGATCATATTGTCTGTTGGCAGATG 5999
Db 683 ysaGluIleLysGlyIleGlnLeuGlyLysGlnGluValLysLeuSerLeuPheAlaAspA 703
QY 6000 ACATGTTTATATTAGAAAACCCATGCTGTACAGGCCAAAACCTCTTAGCTGATTA 6059
Db 703 spMetIleValTyrIleGluGluAsnProIleValSerAlaGlnAsnLeuLeuLysLeuIle 723

QY 6060 GCAACTTCAGCAAAAGTCTCAGAGACACAAAATCAATGTGCAAAAATCAACAGATTCAT 6119
Db 723 xraAsnPheSerLysValSerGlyTyrLysIleAsnValGlnLysSerGlnAlaPheLeu 743
QY 6120 ACCCAATAATATAGCAAAAGAGCCCAATATCATGTGTAACCTCATCAATATGCA 6179
Db 743 yThrAsnAsnArgGlnThrGluSerGlnIleMetLysIleLeuProPheThrIleAlaS 763
QY 6180 CAAGAGAAATTAATACCTAGGAATCAACTTACAGAGGACAGTGAAGACTCTTCAAG 6239
Db 763 erylSarGlyLysTyrLeuGlyIleGlnLeuThrArgAspAlaLysAspLeuPheLysG 783
QY 6240 AGAAGTACAAACCACTGATCAAGAAATAGAGAGACACAAACAAATGAAAACATTC 6299
Db 783 IAsnTyrLysProLeuLeuLysGluIleLysGlnLysPheAsnLysTyrLysAsnIleP 803
QY 6300 CATGCTCACAGATAGTAAGATCAT-----GAAATGCGCATCTAGCCCAAGTAAT 6352
Db 803 rocYSerTyrValGlyArgIleAsnIleValLysMetAlaIleLeuProLysValIleT 823
QY 6353 ATAGATTCAAGTGTACCCCATCAAGCTACATGACTTCTTCCACAGAAATGAAAAA 6412
Db 823 yThrArgPheAsnAlaIleProIleLysLeuProMetThrPheThrGluLeuGluLysT 843
QY 6413 CAACTTAATTTATATATGAAACCAAAAAAGACCCACAGAGCCACAGCAATCTTACG 6472
Db 843 hThrLeuLysPheIleTyrAsnGln-LysArgAlaArgIleAlaLysSerIleLeuSer 862
QY 6473 AAAAAGCAAAAGCTGGAGTATCATGTCACCGATTAACCTATATTAAGGTACA 6532
Db 863 GlnLysAsnLysAlaLysIleThrLeuProAspPheLysLeuTyrLysAlaThr 882
QY 6533 GTAACCAAAACGATGCTACTGCTACCAAAACAGATATATAGACCAATGGAACAGACA 6592
Db 883 ValThrLysThrAlaTyrTyrTyrGlnAsnArgAspIleAspGlnTyrAsnArgThr 902
QY 6593 GAGACCTCGAATATAC-ACGCAATCTCATTCATCTGATCTTTGACAAACCTGACAAA 6651
Db 903 GluProSerGluIleMetProHisIleTyrAsnTyrLeuIlePheAspLysProGlyLys 922
QY 6652 AACCAAGCATGGAAGAAAGATTCCTATTAATTAATGCTGTGGAAGAAACGCGTAGCC 6711
Db 923 AsnLysGlnTyrGlyLysAspSerLeuPheAsnLysTyrCysTyrProLysThrPheAla 942
QY 6712 ATATGCAAGAACTGAAACTGATCCCTTCTTACACCTTATACAAAGTAACTACAGA 6771
Db 943 IleCysThrLysLeuLysLeuAspProPheLeuThrProTyrThrLysIleAsnSerArg 962
QY 6772 TGAATTAAGACTTAATATTAAGACATTAACCAATTAACCA-GAAGAAAACCTAGGC 6830
Db 963 TrpIleLysAspLeuAsnValArgProLysThrIleLysThrLeuGlnGluLysLeuGly 982
QY 6831 AATACCATTCAGATATAGACATGGGCAAAAGCTCATCAATTAACCAACCAAGCAATG 6890
Db 983 IleThrIleGlnAspIleGlyValGlyLysAspPheMetSerLysThrProLysAlaMet 1002
QY 6891 GCAAGAAAACCCAAATATAGACAAGTGGATCTGATTAACATATAGACTTTCACAGCA 6950
Db 1003 AlaThrLysValLysIleAspLysTyrAspLeuIleLysLeuLysSerPheCysThrAla 1022
QY 6951 AAAAAAACTGTATCAGAGTACAGACAGCAACCTACAGATGGAGAAAATTTTGCAT 7010
Db 1023 -LysGluThrThrIleArgValAsnArgGlnProThrThrTyrGlnLysIlePheAlaTh 1042
QY 7011 CTATGATCTGCAAAAGGTATATCCAGAGATCTACAGAAAGCAATTAACAAATTACA 7070
Db 1042 TyrIleSerSerAspLysGlyLeuIleSerArgIleTyrAsnGluLeuLysGlnIleTyrLys 1062
QY 7071 GAAAAA----AACCAACCCCGTCAAAATATAGGCAAGATATAGACAGCACTTCTCAAA 7126
Db 1062 sLysLysThrAsnAsnProIleLysLysTyrAlaLysAspMetAsnArgHisPheSerLys 1082
QY 7127 AGAAGACATTTATGACGCCACAAACATATGAAAAAAACCTCATCATCTTGGCTTAG 7186

```

Db      1082 sgluaspllethryalaalalsylshismetlylscyserserleuallalear 1102
QY      7187 AGAATGCAAAACAAACACAGTGCATACCATCTAGTACTAGTATGATGGATGCAC 7246
Db      1102 gglmetelallethrythrmethlarythhisleuthrprovalargmetalaileil 1122
QY      7247 TAAAGAGTCAGAAACAAACAAATGCTGAGAGAGATGGAGAAATAGAACACTTTTCCA 7306
Db      1122 elyslyssercllysnasnargcystprargcllycysgllylulegylthrmleuulh 1142
QY      7307 CTGTGGTGGAGATGTAATTAAGTTCACACATTTGTGGAAGACAGTGGAGATTCCTTAA 7366
Db      1142 scystprtrpaspysylsleuvalglinproleuthrlysserleuvaltrparphleuvar 1162
QY      7367 GGATCTAGAACCAAAATATCATTTGACCCGACCAATCCCATTTAGTATATCCCAAA 7426
Db      1162 gaspleuengluleuulileprophasprioalalleproleuengllylethrproas 1182
QY      7427 GAATATTAATCATCTCTATTATAAGACACATGCACATATGTTATTGACAGACATGAT 7486
Db      1182 ngulthrlysserlscysyltrlysasprthrCysThrArgmethlevalalaleuaph 1202
QY      7487 CACAATAGCAAAAGCTTGAACCAACCCAAATGTCATCAGTATAGACTGATTAAGAA 7546
Db      1202 ethrillealalythrtrpasinglileuylscysprothrmetileasprtrileuylsly 1222
QY      7547 AACATGGCACATATATACACCAATGATATGATGACCAATAAAG-GATGAGTTCATGTC 7605
Db      1222 smetrhrhislethrythrmethglultrythralalallelysasnspcglupheilese 1242
QY      7606 CTTCGACAGATATGATGATGAGTGAAGACCATCATCTCAGCAACATTAACACAGAAACA 7665
Db      1242 rphevalgllythrtrpethlyslleuengluthrilleleuenserlylseuerglnglul 1262
QY      7666 GAAACCAAAACACCATGTTCTCAGCTGTAAGTGGAGT 7705
Db      1262 nlysthrlyshisarglilepheserleuilegylglsn 1275

RESULT 7
ID      000366 PRELIMINARY; PRT; 1275 AA.
AC      000366
DT      01-JUL-1997 (TREMBLrel. 04, Created)
DT      01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT      01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE      Putative p150.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OX      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE-97285120; PubMed-9140393;
RA      Sassaman D.M., Dombroski B.A., Moran J.V., Kimberland M.L., Naas T.P.,
RA      Debernardinis R.D., Gabriel A., Swergold G.D., Kazazian H.H. Jr.;
RT      "Many human LI elements are capable of retrotransposition."
RL      Nat. Genet. 16:37-43(1997).
DR      EMBL; U93567; AAC51267.1; -.
DR      InterPro; IPR005135; Exo_endo_phos.
DR      InterPro; IPR004777; RVTse.
DR      Pfam; PF03372; Exo_endo_phos; 1.
DR      Pfam; PF00078; rvt; 1.
KW      RNA-directed DNA polymerase.
SQ      SEQUENCE 1275 AA; 149033 MW; 07E88F8F4DB831A2 CRC64;

Alignment Scores:
Pred. No.: 0 Length: 1275
Score: 5182.00 Matches: 1055
Percent Similarity: 78.90% Conservative: 78
Best Local Similarity: 73.47% Mismatches: 137
Query Match: 9.90% Indels: 173
DB: 4 Gaps: 7

```

```

US-10-083-853B-2 (1-29921) x 000366 (1-1275)
QY      3428 ATGGCAGATCAATTCACACATATATATTAATTAACCTTAATGTAATGGCTAAATCC 3487
Db      1 Methrglyssarseniserhislethrilleuuthrleuashilleasnnglyleuanser 20
QY      3488 CCAATTTAAAGACACAGCTGGCAATTTGGATTAAGAGTCAAGCCATGCTGGTGT 3547
Db      21 Alalleysarghisshargleuallaserthrilleysserglinsproserleuvalcyscys 40
QY      3548 ATTGAGAGGCCATCTCACATGAAGAACACACATAGGCTCAAAATTAAGGATGGAGG 3607
Db      41 ileglnghlthrhisleuthrCysargAsprhrhisargleuylsleuylslytrparg 60
QY      3608 AAGATTATCCAGATTAATGGAACCAAAAAAAGAGAGGCTTCATCTCTAGTCTCT 3667
Db      61 LysilethyrglnaalsnglylsGln--LyslyslalaglyValalalleleuvalser 79
QY      3668 GATAAAGACAGCTTTAAACCAACAAAGATCAAAAGAGCAAAAGAGGCCATTAATATG 3727
Db      80 AsplysthrasprleuylsprothrLysilleysargasplysglngllyhstyllmet 99
QY      3728 GTAAAGCATCATATGAACAAAGAGACTAATCTCAATATATATACAGACCCCAATACA 3787
Db      100 Vallysglyserilleglnghlghlghlghlghlghlghlghlghlghlghlghlghlghl 119
QY      3788 GGACACCCAGATTCATTAAGCAAGATTCTTATAGACACTTACACAGACTTTGACTCCAC 3847
Db      120 Glylalproargpreillethysglnvalleuseraspleuenglntargaspleuaspserhis 139
QY      3848 ACATATATAGGGGAGCTTAATTAATATAGACACTTAAACCCCACTGGCCAAATTA 3907
Db      140 Thrleullethetgyl-----AspPheasnthrProleuSerThrleu 153
QY      3908 GGCAGCATATGAGACAGCAAAATTAACAGGATATCCAGAGTGAAGTGAAGTGGTGGAC 3967
Db      154 Hisargserthrarglnlyslvalasnlyslasprhrghlghlghlghlghlghlghlghlghl 173
QY      3968 CAACGGACCTATATAGATATCTACAGAACTCCCAACCCCAATTAACAGAAATATACATC 4027
Db      174 GlnlaalspleuileasprillethryargthrleuhsiprolysSerthrghlghlghlghlghl 193
QY      4028 TTTCACATCATCATCTACACCTATTTAAATGATGACCATGTAATTAAGTAAGAACTGC 4087
Db      194 PheSerlaarprohishsthrtyrSerlyslleasphislevalglyserlyslaleu 213
QY      4088 CTCAGCAATGCAAAAGACAGAAATCTTAACAAACAGTCTTCAGACTACAGTGCATC 4147
Db      214 leuserlyscylsargthlghlillethrsnthyreuserasphissevalale 233
QY      4148 TATTTAGACTCAGAAATTAAGAACTCAGTCAAAATACACACATCACTAGTGAAC 4207
Db      234 Lysleuengluleuarglileysasnleuthlghlserlthrtthrtrplysleuasn 253
QY      4208 AACCTGCTCCGATGACTCTGTTGTAATTAACAAATGAAGGCAAAATTAAGATGTC 4267
Db      254 Asnleuueuasnshprlytrprvalhisasnleuethlyslalagluilleysmerphe 273
QY      4268 TTTGAAACCAATGAGAACAAAGACACAAATGACCAATGTCCTGGGCAATTTAAAGCA 4327
Db      274 PheglunthrsnglunslasprthrThrycglinsnleuthrpaalaphelyala 293
QY      4328 GTGTGTAGAGGGAATTTATAGCTAGATGCCCTACAGAGAAAGCAAGGAATTTCTAAA 4387
Db      294 ValcysargglylsphellealaleuasnlaLarylsarglysglngllylaugSerlys 313
QY      4388 ATAGACACCTTAATCATCAATTAAGAACTAGAGAAAGAGCAAAACAAATTCAAAA 4447
Db      314 Ilesphthrleuthrserghlghlghlghlghlghlghlghlghlghlghlghlghlghlghl 333
QY      4448 GCTAGCAGAAACAGAAATTAATTAAGATCAGACAGACTGAAAGAGATAGACACAA 4507

```


QY 6652 AACAGCAATGAAAAAGATTCCTTAAATTAATTAATGCTGTGGAATAAGCTGATGCC 6711
|||||
Db 923 AsnLysGlnTrpPcylLysAspSerLeuPheAsnLysTrpPcylAsnTrpLeuAla 942
QY 6712 ATTTGCAAGAAAGCTGAAAGCTGGATCCCTTCCTTACACCTTATACAAAAGTTAACTCAAGA 6771
|||||
Db 943 ValCysArgLysLeuLysLeuAspProPheLeuThrProTyrThrLysIleAsnSerArg 962
QY 6772 TGAATTAAGACTTAAATATATAGCAATAAACCATTAATAACCGA-GAAGAAAACCTAGGC 6830
|||||
Db 963 TrpIleLysAspLeuAsnValLysProLysThrIleLysThrLeuGlnAsnLeuGly 982
QY 6831 AATACCATTCAGATATGACATGGCAAGACCTCATGATGATTAACACCAAGCAATG 6890
|||||
Db 983 IleThrIleGlnAspIleGlyValLysAspPheMetSerLysThrProLysAlaMet 1002
QY 6891 GCACAAAAGCCAAAATAGCAAGGAGATCTGATTAACTATAGAGCTTCGACAGCA 6950
|||||
Db 1003 AlaThrLysAspLysIleAspLysTrpAspLeuIleLysLeuLysSerPheCysThrAla 1022
QY 6951 AAAAAAAGCTCATCAGAGTGAACAAGCAAGCAAGCAAGATGGAGAAATTTTGCAT 7010
|||||
Db 1023 -LysGlnThrThrIleArgValAsnArgGlnProThrThrTrpGlnLysIlePheAlaThr 1042
QY 7011 CTATCGATCTGACAAAGCTAATATCCAGATCTACGAGAACTTAACAAATTTACAA 7070
|||||
Db 1042 rTyrSerSerAspLysGlyLeuIleSerArgIleTyrAsnGlnLeuLysGlnIleTyrLys 1062
QY 7071 GAAAAA---AACACCCCGCTCAAAATATGCGCAAGAGATATGACAGACACTTCTCAA 7126
|||||
Db 1062 sLysLysThrAsnAsnProIleLysLysTrpAlaLysAspMetAsnArgHisPheSerLys 1082
QY 7127 AGAAGACATTTATGCGCCCAACAACATATGAAAAAAGCTCATCATCATGCTGCTGCTAG 7186
|||||
Db 1082 sGluAspIleTyrAlaLysLysLysLysCysSerSerLeuAlaIleArg 1102
QY 7187 AGAATGCAAAACCAACACAGCTGATACATACATCATGCTGATTAAGATGGTGCATCAC 7246
|||||
Db 1102 gGluMetGlnIleLysThrThrMetLArgTyrHisLeuThrProValArgMetAlaIleIle 1122
QY 7247 TAAAAAGTCAGGAAAAACAATGCTGGAGAGATGTGGAATAAGAACACTTTTCCA 7306
|||||
Db 1122 eLysLysSerGlyAsnAsnArgCysTrpArgGlyCysGlyLysIleLysThrLeuLeuH 1142
QY 7307 CTGTGATGGAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7366
|||||
Db 1142 sCysTrpTrpAspCysLysLeuValGlnProLeuTrpLysSerValTrpArgPheLeuArg 1162
QY 7367 GGATCTAGAACCCAGAAATATATCATTTGACCCAGCAATCCCATTAAGTATATACCCAAA 7426
|||||
Db 1162 gAspLeuGlnLeuGlnIleLeuProPheAspProAlaIleProLeuLeuGlyIleTyrProLys 1182
QY 7427 GGAATTAATATCTTCTATTATTAAGACACATGACACATATGTTTATTCGACACTGAT 7486
|||||
Db 1182 sAspTyrLysSerCysCysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1202
QY 7487 CAAATAAGCAAGAACTGGAACCAACCAATGTCATGATGATGATGATGATGATGATGATGAT 7546
|||||
Db 1202 eThrIleAlaLysThrTrpAsnGlnProLysCysProThrMetLysLysLysLysLysLysLys 1222
QY 7547 AACATGGCAATATATACCATATGAATATATGAGGCAATTAAG-GATGAGTTCATGTC 7605
|||||
Db 1222 sMetTrpHisIleTyrThrMetGlnLysTrpAlaAlaIleLysAsnAspIleuPheIleLe 1242
QY 7606 CTTTGCAAGATATGATGAGTGGAAACCATCATTTCTGACAAATTAACACAGACACA 7665
|||||
Db 1242 rPheValIleTyrThrMetLysLeuGlnThrIleIleLeuSerLysLeuSerGlnGln 1262
QY 7666 GAAACCAACACACATGTTCTACATTAAGTGGAGAT 7705
|||||
Db 1262 nLysThrLysHisArgIlePheSerLeuIleGlyGlyAsn 1275

RESULT 8

000375
ID 000375 PRELIMINARY: PRT; 1275 AA.
AC 000375:
DT 01-JUL-1997 (TReMBLrel. 04, Created)
DT 01-JUL-1997 (TReMBLrel. 04, Last sequence update)
DE 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Putative p150.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97285120; Pubmed=9140393;
RA Sussman D.M., Dombroski B.A., Moran J.V., Kimberland M.L., Neas T.P.,
RA DeBerardinis R.J., Gabriel A., Swergold G.D., Kazazian H.H. Jr.;
RT "Many human LI elements are capable of retrotransposition.";
RL Nat. Genet. 16:37-43(1997).
DR EMBL; U93572; AAC51276.1; -;
DR InterPro: IPR005135; Exo_endo_phos.
DR InterPro: IPR000477; RVISE.
DR Pfam: PF03372; Exo_endo_phos; 1.
DR Pfam: PF00078; rvt; 1.
KW RNA-directed DNA polymerase.
SQ SEQUENCE 1275 AA; 14905 MW; 456DD0F3DD7A17F5 CRC64;

Alignment Scores:
Pred. No.: 0 Length: 1275
Score: 5182.00 Matches: 1054
Percent Similarity: 79.04% Conservative: 81
Best Local Similarity: 73.40% Mismatches: 135
Query Match: 9.90% Indels: 173
DB: 4 Gaps: 7

US-10-083-853b-2 (1-29921) x 000375 (1-1275)
QY 3428 ATGCGAGCATCAATTCACACATATATATTAATTAATTAATTAATTAATTAATTAATTC 3487
|||||
Db 1 MetThrGlySerAsnSerHisIleThrIleValThrLeuAsnLysGlnLysLeuAsnSer 20
QY 3488 CCAATTAATAAGACACACAGACTGGCAAAATGGATTAAGAGTCAGAACCCATGAGTGTCTG 3547
|||||
Db 21 AlalIleLysArgHisArgLeuAlaSerTrpIleLysSerGlnAspProSerValCysCys 40
QY 3548 ATTCAGAGGCCCATCTCACATGATGAAAAAGACACACATATGCTCAAAATTAAGGATGGAGG 3607
|||||
Db 41 IleGlnGlnThrHisLeuThrCysArgAspThrHisArgLeuLysLysLysGlyTyrArg 60
QY 3608 AAGATTACCCAGTAATGGAACCAACAAAAAGCAAGGGGTGGCAATCTTACTGCTCT 3667
|||||
Db 61 LysIleLysGlnLysLeuThrCysArgAspThrHisArgLeuLysLysLysGlyTyrArg 79
QY 3668 GATTAACACAGACTTTTAACCAACAAAGATCAAAAGAGACAAAGAGCCATTACATTAAG 3727
|||||
Db 80 AspLysThrAspPheLysProThrLysIleLysArgAspLysGlnGlyHisTyrIleMet 99
QY 3728 GTAAGGCATCATGGAACCAAGAGAGCTATCATCTTAATTAATTAATTAATTAATTAATTA 3787
|||||
Db 100 ValLysLysLeuIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 119
QY 3788 GGAGCAAGCCAGATTCATTAAGCAAGTCTTATGAGACCTCAAGAGACTTTGACTGCCAC 3847
|||||
Db 120 GlyAlaProArgPheIleLysGlnValLeuSerAspLeuGlnArgAspLeuAspSerHis 139
QY 3848 ACAATAATAGTGGAGACTTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3907
|||||
Db 140 ThrLeuIleMetGly-----AspPheAsnThrProLeuSerThrLeu 153
QY 3908 GGAGATTCATGAGCAAGAAATTAACAAAGATATCAGAGAGTGAAGAGAGTGGAG 3967
|||||
Db 154 AspArgSerThrArgGlnLysValAsnLysAspThrGlnGlnLysGlnLysGlnLysGlnLys 173
QY 3968 CAAGCGACCTAATATATATCTACAGAACTCCACCCCAAAATCAACAGAAATTAATCACTC 4027

|||||
Db 174 GlnAlaAspLeuIleSerPheIleArgThrLeuHisProLysSerThrGluYrThrPhe 193
QY 4028 TTCTCAGCATCATATACATACCTTTTAAATAGCAGATGATTAATTAAGTAACAACATC 4087
Db 194 PheSerAlaProHisHisThrYrSerLysIleAspHisIleValIleYrSerLysAlaLeu 213
QY 4088 CTCAGCAATTCGCAAAAGAACAGAAATCTTAACAACAAGTCTCTCAGATACAGTCGCAATC 4147
Db 214 LeuSerLysCysLysArgThrGluIleThrAsnTyLeuSerAspHisSerAlaIle 233
QY 4148 TATTTAGACTCAGATTAAGAAACCTCAATCAACAACATCAATGGAAGAACG 4207
Db 234 LysLeuGluLeuArgIleLysAsnLeuSerGlnSerArgSerThrYrPylsLeuAsn 253
QY 4208 AACCTGCTCTGATAGTACTGCTGTAATTAACAAATAGAGCAAAATTAAGATGTTTC 4267
Db 254 AsnLeuTyLeuAsnAspYrTrpValHisAsnGluMetLysAlaGluIleYrMetPhe 273
QY 4268 TTTGAAACCAATGAGAACAAAGACACAAATGTACCAATCTCTGGGCGCATATTTAAGCA 4327
Db 274 PheGluThrAsnGluLysAspThrYrThrGlnAsnLeuTrpAspAlaPheLysAla 293
QY 4328 GTGTGAGAGGGAATTTATAGCACTAGATGCTTAAAGAGAAAGCAGGAATATCTAA 4387
Db 294 ValCysArgGlyLysPheIleAlaLeuAsnAlaTyLysArgLysGlnIleuArgSerLys 313
QY 4388 ATAGACACCTTAACATCACAAATTAAGAAAGTAAGAAAGAAAGCAAAATTCGAAA 4447
Db 314 IleAspThrLeuThrSerIleuLysGluLeuGluLysGlnGluIleThrHisSerLys 333
QY 4448 GCTAGCAGACAGACAGAAATTAACATTAAGATCAGCAGAACTGAGAGATAGACACAA 4507
Db 334 AlaSerArgArgGlnIleuIleThrLysIleArgAlaGluLeuLysGluIleGluThrGln 353
QY 4508 AAAGCCCTCAATTAATGAATGAATCCAGAGCTGTTTGTAAAGATCAAGCAAAAT 4567
Db 354 LysThrLeuGlnLys-MetAsnGluSerArgSerTrpPhePheGluArgIleAsnLysIle 373
QY 4568 AGACCA----CTAGACAGACTATAAAGAAAGAAAGAGAGAAATCAAGATGCAAT 4623
Db 373 AspArgLeuLeuAlaArgLeuIleLysLysLysArgLysAsnGlnIleAspThrIle 393
QY 4624 AAAAAATGATAAGGGGATATCACACCGATCCACAGAAATCAAAATATTATTCAGAGA 4683
Db 393 eLysAsnAspLysGlyAspIleThrThrAspProThrGluIleGlnThrIleArgGlu 413
QY 4684 ATATTATAACACCTATGCAAAATPAACTAGAAAATCAGAAGAAATGATAAATTCCT 4743
Db 413 uTyTrpLysHisLeuTyAlaAsnLysLeuGluLysGluMetAspThrPheIle 433
QY 4744 GGACACATATGTAGCCTGTATGAGCCTGGGGGACAGAACAAAAGGGGGTGAATGCAGAA 4803
Db 433 uAsp----- 434
QY 4804 ATAAAGACAAAGACAAAAGATATGTTTGAAAGTAGGGGTCAAGGGCAACTGCTCT 4863
Db 434 ----- 434
QY 4864 AATGACAAGGGCCCTGAGCTTTACACACACCTCTGTATTTATTAGGCAAAAGAGATAGC 4923
Db 434 ----- 434
QY 4924 GAGAGGTGAGTTGGAAGAAGAGTCAAGCTGTAGGTCCAGAGTAGGCTGCAAGACTGC 4983
Db 434 ----- 434
QY 4984 ATTCTCAAAACAATAGGCTCTAGATGTCCCACTAGATTAACCTCAAGAGACCGAGTCCAGG 5043
Db 434 ----- 434
QY 5044 GAGTATGGCCCTCAGCAAAACCTCTTAGGGGACGACAGAAAGTAAGTTTGCCACATTCCT 5103

Db 434 ----- 434
QY 5104 GATATCCAGATAAACAGATTGGCTGTGTGATCAAGTAGACCTCCAGTGAATGCTGAGTTGG 5163
Db 434 ----- 434
QY 5164 TCATGATCCCTTTGGCCTTTTGGCTCCCAAAACATACACCTCTCAAGCTAAACCA 5223
Db 435 -----ThrTyThrLeuProArgLeuAsnGlu 443
QY 5224 GGAGAGACGCAATCCCGTAAATATACCAATGCAAGTCTAAATTAAGACGTAATGCA 5283
Db 443 nGluGluValGluSerLeuAsnArgProIleThrGlySerGluIleValAlaIleIleAs 463
QY 5284 TAGCCTACCAACCAAAAAAGTCCAGGACAGACGATTCACAGCCAAATTTACAGAG 5343
Db 463 nSerLeuProThrLysLysSerProGlyProAspGlyPheThrAlaGluPheTyGlnArg 483
QY 5344 GTACAAAGAGAGCTGGTACTATTCCTCTGAAACTATTCCAAAAATRGAA---AATGG 5400
Db 483 gTyTrpGluGluLeuValProPheLeuLysLeuPheGlnSerIleGluLysGlu-G 503
QY 5401 GAATCCCTCACTCATTTTACGAGGCCAGCATCATCTGATACCAAAACCTAGACAGT 5460
Db 503 LylleuProAsnSerPheTyGluAlaSerIleIleLeuIleProLysProGlyArg 523
QY 5461 ACACAACAAAAGAGAAATTCAGGCCCATATCCCTGATGCAATGTGATGAAATTC 5520
Db 523 sPTThrLysLysGluAsnPheArgProIleSerLeuMetAsnIleAspAlaLysIle 543
QY 5521 TCAATTAATTAATGTGGCAAAATTCAGACACACATCAAAAAGCTTATCTACATGATC 5580
Db 543 euAsnLysIleLeuAlaAsnArgIleGlnGlnHisIleLysLysLeuIleHisAspG 563
QY 5581 AAGTGGCGTCATCCCTGGATGTCAGAGCTGGTCAAAATTAAGCAAAATTAATGTG 5640
Db 563 InValGlyPheIleProGlyMetGlnGlyTrpPheAsnIleHisLysSerIleAsnValI 583
QY 5641 GCCATCACATAAAGCAAGCAATGACAAAAACACATGATTTATCTCATAGTAGCAGAA 5700
Db 583 leGlnHisIleAsnArgAlaLysAspLysAsnHisMetIleIleSerIleAspAlaGlu 603
QY 5701 AGGCTTTGTCAAAATTCAGACGCCCTTCATCTTAATAATTCAGTAACATGATTCG 5760
Db 603 yAlaPheAspLysIleGlnGlnProPheMetLeuLysThrLeuAsnLysLeuGlyIleA 623
QY 5761 ATGGAATGATCTCAAAATTAATAGAGCATTTATAC--AAACCCAGACCAATATCATAC 5819
Db 623 sPGLyThrTyPheLysIleIleArgAlaIleTyAspLysProThrAlaAsnIleIle 643
QY 5820 TGAATGGGCAAAACTGGAAGCATTCCTTTGAGAACTGGCACAAGAGAGATGCCCTC 5879
Db 643 euAsnGlyLysLysLeuGluAlaPheProLeuTyThrGlyThrArgGlnGlyCysPro 663
QY 5880 TTCTACACACTCTATTCAGAGATACATTTGGAAGTTCTGGCCGCGCATCAGGCAATAGA 5939
Db 663 euSerProLeuLeuPheAsnIleValLeuGluValLeuAlaArgAlaIleArgGlnGlu 683
QY 5940 AAGAAATAAAGGATTTCAATATAGAAAGAGAGAGATCATATTTCTGTGTCAGAGT 5999
Db 683 ySGluIleLysGlyIleGlnLeuGlyLysGluGluValLysLeuSerLeuPheAlaAsp 703
QY 6000 ACATGTTGTATATTAGAAAACCCATCGTCTCAGGCCAAAACCTCTTAAGCTGATPA 6059
Db 703 sPmetIleValTyLeuGluAsnProIleValSerAlaGlnAsnLeuLeuLys 723
QY 6060 GCACTTCAGCAAAATTCAGACACAAAATCAATGTCGCAAAATCACAAGATTCAT 6119
Db 723 eAsnPheSerLysValSerGlyTyLysIleAsnValGlnLysSerGlnAlaPheLeuT 743
QY 6120 ACGCAATATATGACAAACAGAGACCAAAATCATAGAGAACTCATTCATCAATTGCTA 6179
Db 743 yThrAsnAsnArgGlnThrGlnSerGlnIleMetSerGluLeuProPheThrIleAla 763

```

QY 6180 CAAGAGAAATAAATACCTAGCAATCACTTACAAAGGACACAGTAGCACTCTCAGG 6239
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 763 eIrysaTgIleYsTyrleuGlyLleGlnleuthrAsgaPValysAsPleuPhelysG 763
QY 6240 AGAAGTACAAACACCTGATCAGGAATTAAGAGAGACACAAACAAATGGAAAAACATTC 6299
   lAsnTyTlysProleuLleuYsgIuileYsgIuAspThrAsnTyTlysAsnIleP 803
Db 783 lAsnTyTlysProleuLleuYsgIuileYsgIuAspThrAsnTyTlysAsnIleP 803
QY 6300 CATGCTCACAGATGATGAAGATCAT-----GAAATGCCATCTGCCAAAGTAAT 6352
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 803 rOcysSerTrpValGlyArgIleAsnIleValysMetAlaIleleuProLysValIleT 823
QY 6353 ATAGATTTCAGTGTACCCCATCAGCTACAGTGTGACTTCTTCACAGATTTGAAAAA 6412
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 823 yTrArgPheAsnAlaIleProIleYsleuPromeThrPhePheThrGlnleuGluYst 843
QY 6413 CAACCTTAAATTCATATGGAACCAAAAAAGCCACAGACCACAGACATCTTAAGC 6472
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 843 hrThrleuLysPheIleTrpAsnGln-LysArgAlaArgIleAlaIysSerIleLeuSer 862
QY 6473 AAAAAGAACAAAGAGTATACCTACCTACCTGACTTAATAACTATATCTTAAGGCTACA 6532
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 863 GlnLysAsnLysAlaGlyGlyIleThrLeuProAspPheLysleuTyrlsLysAlaThr 882
QY 6533 GTAACCAAAACCTGATGCTACTGCTACCAAAACAGATATATAGACCAATGCAAGACA 6592
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 883 ValThrlYsThrAlaTrpTyrrTrpTyrglnAsnArgAspIleAspGlnTrpAsnArgThr 902
QY 6593 GAGACCTCAGAAATAC-ACAGCAATCTACATCATCTGATCTTGAACCAACCTGACAAA 6651
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 903 GluProSerGlnIleMetProHisIleTyrlAsnTyrlleuIlePheAspLysProGlnLys 922
QY 6652 AACACCAATGGAAGAAAGATTCCTATTAATTAAGGCTGGAAAACTGGCTAGCC 6711
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 923 AsnLysGlnTrpLysAspSerLeuPheAsnLysTrpCysTrpGlnAsnTrpLeuAla 942
QY 6712 ATATGAGAAAGCTGAAAGCTGATCCCTCTACACCTATATCAAAAGTAACTACTAGA 6771
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 943 IleCysArgLysleuLysleuAspProPheleuThrProTyrlThrlYsIleAsnSerArg 962
QY 6772 TGAATTAAGACTTAATATATAGACATTAACCATAAACCA-GAAGAAACCTAGGC 6830
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 963 TrpIleLysAspLeuAsnValLysProLysThrIleYsThrleuGlnGluAsnleuGly 982
QY 6831 AATACCATTCAGATATGACATGGGCAAGACTTCTATGCTAAACACCAAAAGCAATG 6880
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 983 IleThrlleGlnAspIleGlyValGlyLysAspPheMetSerLysThrProLysAlaMet 1002
QY 6891 GCAACAAAACCCAAATAGACAAGTGGATCTGATTAACCTATAGACTCTCGACAGACA 6950
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1003 AlaThrlYsAlaLysIleAspLysTrpAspLeuIleLysleuYsSerPheCysThrAla 1022
QY 6951 AAAAAAACTGTCATCAGAGTGAACAACAACCTACAGATGGAGAAAAATTTTGGCAAT 7010
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1023 -LysGlnThrThrIleArgValAsnArgInProThrThrTrpGlnLysIlePheAlaSe 1042
QY 7011 CTATCGATCTGCAAAAGGCTTAATATCAGAGATCTACGAAGACTTAACAAATTTTCAA 7070
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1042 rTyrlSerSerAspLysGlyLeuIleSerArgIleTyrlAsnGlnleuYsglnIleTyrl 1062
QY 7071 GAAAAA-----AACACCCCGCTCAAAATATGGCAAGATGATGAGACACTTCACAA 7126
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1062 sTyrlsTyrlAsnAsnProIleLysTyrlPalalYsAspMetAsnTrgHisPheSerIly 1082
QY 7127 AGAAGACATTTATGACAGCCAAACAACATATGAAAAAACCTCATCATCTGCTGCTAG 7186
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1082 sGlnAspIleTyrlAlaThrLysIlyshMetLysCysSerSerleuAlaIleAr 1102
QY 7187 AGAAATGCAAAACAAACACAGTGCATATCCATCTCATCTGCTGATGATGATC 7246
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1102 gGlnMetGlnIleLysThrThrMetArgTyrlHisLeuThrProValArgMetAlaIleIl 1122

```

```

QY 7247 TAAAACTCAGAAACAAACAAATCTGAGAGAGATGTGGAATATAGAACTTTTCCA 7306
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1122 eLysLysSerGlyYAsnAsnArgCysTrpArgGlyCysGlyGlnIleGlyThrleuHl 1142
QY 7307 CTGTGGTGGGAATGTAATTAATTAAGTCAACATTTGTGAGACAGCTGTGAGATTCCTTAA 7366
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1142 sCysTrpTrpAspCysLysleuValGlnProleuTrpLysSerValTrpArgPheLeuAr 1162
QY 7367 GGAATCTGAACCAAGAAATATCATCTTGACCCAGCAACCCCAATTCAGTATATACCCAAA 7426
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1162 gAspLeuGlnleuGlnIleProPheAspProAlaIleProleuGlnIleTyrlProAs 1182
QY 7427 GGAATTAATTAATCATCTATTAATTAAGACACATGACACATATGTTATTCACAGACTGAT 7486
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1182 ngIuTyrlYsSerCysCysTyrlYsAspTrnCysThrArgMetPheIleAlaAlaLeuPh 1202
QY 7487 CACAAATGCAAAAGACTTGGAAACCAACCCAAATGTCATCAGTGTAGACTGTAAAGAA 7546
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1202 eThrlleAlaLysThrTrpAsnGlnProLysCysProThMetIleAspTrpIleLysly 1222
QY 7547 AACATGGCAGATATACACCCAAATGAAATCTATGCGACCATTAAGG-GATGAGTTCATGTC 7605
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1222 sMetTrpHisIleTyrlThrMetGluTyrlYrlAlaAlaIleLysAsnAspGluPheIleSe 1242
QY 7606 CTTTGCAGAGATATGATGATGAGAGCTGGAACCATCATCTGACGAACCTAACACAAAGACA 7665
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1242 rPheValGlyThrTrpMetLysleuGlnThrlleIleLeuSerLysleuSerGlnGluG 1262
QY 7666 GAAACCAAAACCCACATGTTCTCAGTTGTGAAGTGGAGT 7705
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1262 nLysThrLysHisArgIlePheSerleuIleGlyGlyAsn 1275

```

RESULT 9

```

ID Q15606 PRELIMINARY: PRT: 1275 AA.
AC Q15606:
DT 01-NOV-1996 (TREMblrel..01, Created)
DT 01-NOV-1996 (TREMblrel..01, Last sequence update)
DT 01-JUN-2002 (TREMblrel..21, Last annotation update)
DE Hypothetical 149.0 kDa protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Dombroski B.A.;
RL Submitted (JAN-1992) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 84-1275 FROM N.A.
RA Woods-Samuels P.;
RL Submitted (APR-1989) to the EMBL/GenBank/DBJ databases.
DR EMBL: M80343; AAB59368.1; -.
DR EMBL: M22333; AAB88037.1; -.
DR InterPro: IPR005135; Exo_endo_phos.
DR InterPro: IPR000477; RVTse.
DR Pfam: PF03372; Exo_endo_phos; 1.
DR Pfam: PF00078; tvc; 1.
KW Hypothetical protein; RNA-directed DNA polymerase.
SQ SEQUENCE 1275 AA; 148971 MW; 6DA1229742848323 CRC64;

```

Alignment Scores:

```

Pred. No.: 0 Length: 1275
Score: 5176.00 Matches: 1054
Percent Similarity: 78.83% Conservative: 78
Best Local Similarity: 73.40% Mismatches: 138
Query Match: 9.89% Indels: 173
DB: 4 Gaps: 7

```

US-10-083-853b-2 (1-29921) x Q15606 (1-1275)

```

QY 3428 ATGGCAGGATCAAAATTCACACATTAATTAATTAACCTTAATGTAATGGGCTAAATTC 3487
   ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

Db 1 MetThrGlySerAsnSerHisIleThrIleLeuThrIleuAsnIleAsnGlyLeuAsnSer 20
QY 3488 CCAATTTAAAGACACAGACTGGCAATTTGGATTAAGAGTCAGACCCATCAGTGTGCT 3547
Db 21 AlaIleLysArgHisIleuAlaSerTrpIleLysSerGlnAspProSerValLysCys 40
QY 3548 ATTACAGAGGCCCATCTGCATCATGAAAAAGACACACATAGGCTCAAAATTAAGGATGGAGC 3607
Db 41 IleGlnGluThrHisIleuThrCysArgAspThrHisIleuLysIleLysGlyTrpArg 60
QY 3608 AAGATTACCAAGTAATGGAAAAACAAAAAAGAGAGGAGGCTTCAATCCAGTCTCT 3667
Db 61 LysIleTyrGlnAlaAsnGlyLysGln---LysLysAlaGlyValAlaIleLeuValSer 79
QY 3668 GATAAACAAGACTTTTAACCAACAAGATCAAAAGACAAAGAGAGCCATTACATATG 3727
Db 80 AspLysThrAspPheLysProThrLysIleLysArgAspLysGluLysIleuLysTyrIleMet 99
QY 3728 GTAAGAGCATCATGACAGAAAGAGAGCTACATCTAATATACATGACACCCATATCA 3787
Db 100 ValLysGlySerIleGlnGlnGluLeuThrIleuAsnIleTyrAlaProAsnThr 119
QY 3788 GGAGCACCAGATTCATTAAGCAAGTTCTTAGAGACTACAAAGAGACTTTGACTCCAC 3847
Db 120 GlyAlaProArgPheIleLysGlnValLeuSerAspLeuGlnArgLysLeuAspSerHis 139
QY 3848 ACATATATAGTGGAGACTTAATATTAATAGACACTTTACACCCCACTGCCAATATTA 3907
Db 140 ThrLeuIleMetLys-----AspPheAsnThrProLeuSerThrIle 153
QY 3908 GGCAGATCATGACAGAAATTAATACAGGATATCCAGAGTGAAGCTGAGCTGGAC 3967
Db 154 AspArgSerThrArgGlnLysValAsnLysAspThrGlnGluLeuAsnSerAlaLeuHis 173
QY 3968 CAAGCGACCTATATAGATATCTACAGAACTCCGCCCAATCAACAGATATACATC 4027
Db 174 GlnAlaAspLeuIleAspIleTyrArgThrIleuHisProLysSerThrGlyTrpIlePhe 193
QY 4028 TTCTCGACATCACAATTACACCTATTTAAATGACCAATGTAATTTAAAGTAACAACCTC 4087
Db 194 PheSerAlaProHisIleThrTyrSerLysIleAspHisIleValLysSerLysAlaLeu 213
QY 4088 CTCAGCAATGCAAAAGACAGAAATCTTAACAAGTCTCAGACTACATGACATGCAATC 4147
Db 214 LeuSerLysCysLysArgThrGluIleThrHisTyrLeuSerAspHisSerAlaIle 223
QY 4148 TATTTAGAATCAGAAATTAAGAAATCTACCTCAAAATACACAACTACATGGAAGTGAAC 4207
Db 234 LysLeuGluLeuArgLysIleLysAsnLeuThrGlnSerArgSerThrTrpLysLeuAsn 253
QY 4208 AACCTGCTCTGATGATGACTACCTACCTGCTAATTAACAATGAGCAAAATTAAGATGTTT 4267
Db 254 AsnLeuLeuLeuAsnAspTyrTrpValHisAsnGluMetLysAlaIleLysMetPhe 273
QY 4268 TTGTAACCAATGAGAAACAAGACACATGATACCAATCTCTGGGCACTATTATTAAGCA 4327
Db 274 PheGluThrAsnGluAsnLysAspThrTrpTyrGlnAsnLeuTrpAspAlaPheLysAla 293
QY 4328 GTGTGTAGAGGAAATTTATAGCACTAGATGCTTACAGAGAAAGAGAGCAAAATATATTA 4387
Db 294 ValLysArgGlyLysPheIleAlaLeuAsnAlaTyrLysArgLysGlnGluLeuArgSerLys 313
QY 4388 ATAGACACCTTAAATCACAATTTAAAGAACTTAGAGAGAAAGAGCAAAATTAATTA 4447
Db 314 IleAspThrLeuThrSerGlnLeuLysGluLeuGlnGluLeuGlnIleThrHisSerLys 333
QY 4448 GCTACAGAGAGCAAGAAATTAATACTAGATCAGAGCAAGCACTGAAGAGATAGAGACACA 4507
Db 334 AlaSerArgArgGlnGluIleThrLysIleArgAlaGluLeuLysGluIleGluThrGln 353
QY 4508 AAAGCCCTCAATTAATCAATGAATCCAGAGCTGTTTGTGAAAGAGACAGCAAAAT 4567
Db 354 LysThrLeuGlnLys-IleAsnGluSerLysSerTrpPheGluArgIleAsnLysIle 373

QY 4568 ----AGACCACTAGACAGACTAATTAAGAAAGAGAGAGAAATCAAGATCAAT 4623
Db 373 eAspArgProLeuAlaArgLeuIleLysLysLysArgLysLysAsnGlnIleAspThrIle 393
QY 4624 AAAAATGATTAAGGGGATATCAACCCAGTCCACAGAAATCAAACTTTATTCAGAG 4683
Db 393 eLysAsnAspLysGlyAspIleThrThrAspProThrGluIleGlnThrIleArgGlu 413
QY 4684 ATATTATTAACACCTCTATGCAATTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTCT 4743
Db 413 uTyrTyrLysHisLeuTyrAlaAsnLysLeuGluAsnLeuGlnGluMetAspThrPheLe 433
QY 4744 GACACATATATGACCTGTATGACCTTTGGGGACAGAACAAAGGGGATGAATGCAGAA 4803
Db 433 uAsp----- 434
QY 4804 ATAAAGACAAAGACAAAGAGATATGTTGGAAGTAGGGGTACAGGGGCAATTCGCTCT 4863
Db 434 ----- 434
QY 4864 AATGACAAGGGCCCTGAGCTTTACACACCCCTGTATTTATTAAGCAAAAGAGATAGC 4923
Db 434 ----- 434
QY 4924 GAGAGGAGTGAAGTGAAGAGAGAGTCAAGCTGTAGTCCAGAGTAGGCTCGAAGACTGC 4983
Db 434 ----- 434
QY 4984 ATTCTCAAAACANTAGGCTCTAGATGTCACAGTAGATTAACCTCAAGAGCAGTGCAGG 5043
Db 434 ----- 434
QY 5044 GAGTATGAGCCCTCAGCAAACTCTTAGAGGACAGACAGAAATGATTGCCACATCTCT 5103
Db 434 ----- 434
QY 5104 GTATTACGATTAACAGTTTGCTGTTCATCAAGTAGCTCCAGTGGATGCTGAGTGG 5163
Db 434 ----- 434
QY 5164 TCATGATCCCTTTGGCTTTTGGTCCCAAAACATACATACCCCTCAAGACTAAACA 5223
Db 435 -----ThyTrpThrLeuProArgLeuAsnGlu 443
QY 5224 GGAAGAGTCGAAATCCCTGAATATACCACTACCACTCTTAATTAAGACGTAATGA 5283
Db 443 nGluGluValGluSerLeuAsnArgProIleThrGlySerGluIleValAlaIleIleAs 463
QY 5284 TAGCCTACCAACCAAAAGATCCAGACAGAGGATTCACAGCCAAATTCATACAGAG 5343
Db 463 nSerLeuProThrLysLysSerProGlyProAspGlyPheThrAlaGluPheTyrGlnArg 483
QY 5344 GTACAAAGAGAGCTGTACTATTCCTTGTGAACATATCCAAAAATAGAA---AATG 5400
Db 483 gTyrMetGluGluLeuValProPheLeuLeuLysLeuPheGlnSerIleGluLysGlu-6 503
QY 5401 GAATCTCTCTAATCATTTTACAGAGCAGCATCTGATACCAAAACCTAGCAGTG 5460
Db 503 lYleLeuProAsnSerPheTyrGluAlaSerIleIleLeuIleProLysProGlyArgA 523
QY 5461 ACACAACAAAGAGAAATTTTCAGGCCCATATCCCTGATGAACATTTGATGTGAATAATC 5520
Db 523 sPThrThrLysLysGluAsnPheArgProIleSerLeuMetAsnIleAspAlaLysIle 543
QY 5521 TCAATTAATTAATGCGCAACCAATTCACAGACATCAAAAGCTTATCTACCAAGATC 5580
Db 543 euAsnLysIleLeuAlaAsnArgIleGlnGlnHisIleLysLysLeuIleHisAspG 563
QY 5581 AAGTTGGCTCATCCCTGGATGCAAGGCTGTGTTCAAAATATGCAATCAATCAATATGTG 5640
Db 563 InvalGlyPheIleProGlyMetGlnGlyTrpPheAsnIleArgLysSerIleAsnValI 583

QY 5641 GCCATCATATAACAGAACCAATGACAAAAACCATGATATATTCATATGACAGAAA 5700
|||||
Db 583 lEGLnHlIleAsnAtgAlaLysAspLysAsnHlIleMetIleSerIleAspAlaGlu 603
QY 5701 AGGCCTTTCGAAATATCAACAGCCCTTCATGCTAAATTCCTCACTAACTATGATG 5760
|||||
Db 603 yAlaPheAspLysIleGlnGlnProPheMetLeuLysThrLeuAsnLysLeuGlyLea 623
QY 5761 ATGGAATGTATCTCAAAATATATAGAGCATTTATAC-AAACCCACAGCCATATATAC 5819
|||||
Db 623 spGlyThrThrPheLysIleIleArgAlaIleTyrAspLysProThrAlaAsnIleIle 643
QY 5820 TGAATGGCCAAAACTGAGAGCATTCCTTTGAGAACTGGCACAGACAGATGGCCCTC 5879
|||||
Db 643 euAsnGlyGlnLysLeuGlnAlaPheProLeuLysThrGlyThrArgGlnGlyCysPro 663
QY 5880 TCTCACACTCTTATCAAGATATCTTGAAGTTCTGGCCAGGCAATCAGCAATYAGA 5939
|||||
Db 663 euSerProLeuLeuPheAsnIleValLeuGlnValLeuAlaArgAlaIleArgGlnGlu 683
QY 5940 AAGAATAAAGGTATTCATATAGAAAGAGAGAGCAATCATTTGCTCCTGTTGACAGT 5999
|||||
Db 683 yGlnIleLysGlyIleArgLeuGlyLysGlnGlnValLysLeuSerLeuPheAlaAsp 703
QY 6000 ACATGTTTGTATATTAGAAAAACCCATGCTCTCAGGCCAAAACTCTTAAGCTGATA 6059
|||||
Db 703 spMetIleValIleTyrLeuGlnAsnProIleValSerIleGlnAsnLeuLeuLysLeu 723
QY 6060 GCAACTTCGCAAAAGTCTCAGACACACAAATCAATGSCAAAAATACACAGCATTTCT 6119
|||||
Db 723 eTAsnPheSerLysValSerGlyTyrLysIleAsnValGlnLysSerGlnAlaPheLeu 743
QY 6120 AGCCCAATATATGACAAACAGAGAGCCCAATCATGCTGATCATCATACATGCTA 6179
|||||
Db 743 yThrAsnAsnArgGlnThrGlnLysSerGlnIleMetGlyLysLeuProPheValIleAs 763
QY 6180 CAAGAAGATATAATACCTAGAGATCAACTTACAGAGGACAGTGAAGACTCTTCAAG 6239
:::|||||
Db 763 eTyrAsnArgIleLysTyrLeuGlyIleGlnLeuThrArgAspAlaLysAspLeuPheLys 783
QY 6240 AAGACTACAACCACTGATCAAGGAATATAGAGAGACACAAACAAATGAAAAACATTC 6299
|||||
Db 783 lAsnTyrLysProLeuLeuLysGlnIleLysGlnLysPThrAsnLysTyrLysAsnIleP 803
QY 6300 CATGCGACAGATAGATGATCAT-----GAAATGCAATACAGCCCAAGTAAAT 6352
|||||
Db 803 rCysSerTyrAlaGlyAlaIleAsnIleValLysMetAlaIleLeuProLysValIle 823
QY 6353 ATGATTCAGTCTACCCCATCAACCTACATTCATCTTCTTCACAGATTTGAAAAA 6412
|||||
Db 823 yArgPheAsnAlaIleProIleLysLeuProMetThrPheThrGlnLeuGlnLys 843
QY 6413 CAACCTTAATTTTCATATGGAACCAAAAAAGAGCCACAGAGCCCAAGCAATCTTAAG 6472
|||||
Db 843 hThrLeuLysPheIleTyrAsnGln-LysArgAlaArgIleAlaLysSerIleLeuSer 862
QY 6473 AAAAAGAACAAAGCTGAGTATCATCTGACCTGATTAATACTATACATTAAGGTACA 6532
:::|||||
Db 863 GlnLysAsnLysAlaGlyIleThrLeuProAspPheLysLeuTyrTyrLysAlaThr 882
QY 6533 GTAACCAAAAGTGCATGCTAGTGTACCAAAACAGATATAGACCAATGGAACAGAGAC 6592
|||||
Db 883 ValThrLysThrAlaTyrTyrTyrGlnAsnArgAspIleAspIleThrAsnArgThr 902
QY 6593 GAGACCTCGAATATAC-ACGCAATCTACATCCATCTGATCTTTGACAAACCTGACAA 6651
|||||
Db 903 GluProSerGlnIleMetProHisIleTyrAsnTyrLysIlePheAspLysProGlnLys 922
QY 6653 AACCAAGCAATGAAAAAGATTCCTATTATTAATGTTGAAAAAGCGCTAGGC 6711
|||||
Db 923 AsnLysGlnTyrIleLysAspSerLeuPheAsnLysTyrCysTyrIleAsnThrIleAla 942
QY 6712 ATATGCAAGAAAGCTGAAGCTGATCCCTTCTTACACCTTATACAAAGTTAACTACAG 6771

Db 943 lIleCysArgLysLeuLysLeuAspProPheLeuThrProTyrThrLysIleAsnSerArg 962
QY 6772 TGAATTAAGACTTAATATTAAGACATTAACCATTAACCAACCA-CAAGAAAACTTAGC 6830
|||||
Db 963 TrpIleLysAspLeuAsnValLysProLysThrIleLysThrLeuGlnGlnLeuGly 982
QY 6831 AATACCATTCAGCATATGACATGAGCGCAAAAGCTTCATACCTTAACCAACCAAGCAAG 6890
|||||
Db 983 lIleThrIleGlnAspIleGlyValIleLysAspPheMetSerLysThrProLysAlaMet 1002
QY 6891 GCAACAAAAAGCCAAATATAGACAGTGGATCTGATTAACTATAGAGCTTCTGACAGCA 6950
|||||
Db 1003 AlaThrLysAspLysIleAspLysTyrAspLeuIleLysLeuLysSerPheCysThrAla 1022
QY 6951 AAAAAAAGCTGCATCAGAGTGCACAGACAGCACTACAGATGGGAGAAAAATTTTGCAAT 7010
|||||
Db 1023 -LysGlnThrThrIleArgValaAsnArgGlnProThrThrTrpGlnLysIlePheAlaThr 1042
QY 7011 CTATGCATCTGACAAAGCTAATATCCAGAGATCTACGAAAGCACTTAACCAATTTACAA 7070
|||||
Db 1042 rTyrSerSerAspLysGlyLeuIleSerArgIleTyrAsnGlnLeuLysGlnIleTyr 1062
QY 7071 GAAAAA-----AACCAACCCGCTCAAAATATGGCCAAAGATATAGACAGACACTTCAAA 7126
|||||
Db 1062 sLysLysThrAsnAsnProIleLysLysTrpAlaLysAspMetAlaAsnArgHisPheSer 1082
QY 7127 AGAAGCATTTATGACGCCACAAACATATGAAAAAAGCTCATCATGTCGCTAG 7186
|||||
Db 1082 sGlnAspIleTyrAlaAlaLysLysHisMetLysCysSerSerSerLeuAlaIleArg 1102
QY 7187 AGAATGCAAAACAAACACAGTACATACCTTCATGCTCAGTATGAAATGGTGCATAC 7246
|||||
Db 1102 gGlnMetGlnIleLysThrThrMetArgTyrHisLeuThrProValArgMetAlaIle 1122
QY 7247 TAAAAAGTCAGAAACAAACAAATGCTGAGAGAGATGGAATATGGAACACTTTTCCA 7306
|||||
Db 1122 eLysLysSerGlyAsnAsnArgCysTyrTrpArgLysGlyGlnIleGlyThrLeuLeuH 1142
QY 7307 CTGTGCTGGGAATGTAATTAATTAATTCATCATCTGTGGAAGACAGTGTGAGATTCCTTAA 7366
|||||
Db 1142 sCysTrpTyrPaspCysLysLeuValGlnProLeuThrLysSerValTyrTrpPheLeuArg 1162
QY 7367 GCACTTAGAACCGAATATATCATTTGACCCAGCAATCCCATTTCTGAGTATTAACCAAA 7426
:::|||||
Db 1162 GasPheGlnLeuGlnIleProPheAspProAlaIleProLeuLeuGlyIleTyrProAs 1182
QY 7427 GGAATATTAATCTTATATTAAGACACATGACACATATGTTTATGACAGCTGAT 7486
|||||
Db 1182 nGlnTyrLysSerCysCysTyrLysAspThrCysThrArgMetPheIleAlaAlaLeuP 1202
QY 7487 CACAATGCAAGACTGTGACCAACCAACCAATGTCATGATGATGATGATGATTAAGAA 7546
|||||
Db 1202 eThrIleAlaLysThrTrpAsnGlnProLysCysProThrMetIleAspTrpIleLys 1222
QY 7547 AACATGCAATATACCATGATGAATATCATGAGCCATAAAAAG-GATGAGTTCATGTC 7605
|||||
Db 1222 smetTrpHisIleTyrThrMetCylutyrTyrAlaAlaIleLysAsnAspGlnPheLeu 1242
QY 7606 CTTTGCAAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7665
|||||
Db 7666 GAAACCAACCAACCAATGCTTCATCTGATGATGATGATGATGATGATGATGATGAT 7705
|||||
Db 1262 nLysThrLysHisArgIlePheSerLeuIleGlyLysAsn 1275

RESULT 10
ID 000363 PRELIMINARY; PRT: 1275 AA.
AC 000363;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, last sequence update)


```
QY 5164 TCATGATCCCTTTGGCTTTTGGCTCCCAAAACATACACCTCTCAAGATAAACA 5223
Db 435 -----ThrTyrThrLeuProArgLeuasnG1 443
QY 5224 GGAAGAAGTCAAATCCCTGATATACCAAGTCTAAATGAGCACTAATGTA 5283
Db 443 ngIuGluValG1SerLeuasnArgProIleThrG1SerGluIleValAlaIleIleas 463
QY 5284 TACCCTACCAACAAAAAAGTCCAGACGACGAGATTCACGCCAAATTCACAGAG 5343
Db 463 nSerLeuProThrLysSerProG1yProAspG1yPheThrAlaG1yPheTyrGlnAr 483
QY 5344 GTCAAGAAGAGAGCTGTACTATTCTTGTGAAGTATTCACAAAATAAGAA---AATGG 5400
Db 483 gTyrLysG1uGluLeuValProPheLeuLeuLysLeuPheGlnSerIleG1uLysGlu-G 503
QY 5401 GAATCTCCCTAATCATTTTACGAGGCCAGCATCATCTGATACCAAAACCTAGCAGTG 5460
Db 503 LylLeuProAsnSerPheTyrG1uAlaSerIleIleLeuIleProLysProG1yArgA 523
QY 5461 ACCAACAAAAAGAGAAATTTTCAGGCCCATATCCCTGATGAGCACTTGAATGTAATCC 5520
Db 523 sPthrThrLysLysG1uAsnPheArgProIleSerLeuMetAsnIleAspAlaLysIleL 543
QY 5521 TCATATAAATCTGCGAAACCAATCCAGCAGACATCAAAAAGCTTATCTACATGATC 5580
Db 543 euAsnLysIleLeuAlaAsnArgIleGlnGlnIleHisIleLysLysLeuIleHisIleAspG 563
QY 5581 AACTGGCGTCATCCCTGGGATGCAAGGCTGTCAAAATATGCAAAATCAATTAATGTAG 5640
Db 563 InValG1yPheIleProG1yMetGlnG1yTyrPheAsnIleArgLysSerIleAsnValI 583
QY 5641 GCATACATTAACCAACCAATGATGATATCTCAATTCATGATGACAGAAA 5700
Db 583 legIlnIleAsnAlaGalaLysAspLysAsnIleMetIleIleSerIleAspAlaGluL 603
QY 5701 AGGCTTTTGTCAAAATTCACACGCTTCGTAAAAATCTCAGTAAACTAGTATCG 5760
Db 603 ysaIaPheAspLysIleGlnGlnProPheMetLeuLysThrLeuAsnLysLeuG1yIleA 623
QY 5761 ATGGAATGTATCTCAAAATATTAAGACTATTTATAC-AAACCACACCAATATCATAC 5819
Db 623 spE1yMetLysrPheLysIleIleIleArgAlaIleTyrAspLysProThrAlaAsnIleIleL 643
QY 5820 TGAATGGCAAAACCTGGAAGCATTCCTTGTGAGACTGGCACAAGACAGATGCCCTC 5879
Db 643 euAsnLysGlnLysLeuGluAlaPheProLeuLysThrG1yThrArgGlnG1yCysProL 663
QY 5880 TCTCACCACTCTATTCAGATACTATTGGAAGTCTGGCCAGGCAATCAGCAATAGA 5939
Db 663 euSerProLeuLeuPheAsnIleValLeuGluValLeuAlaArgAlaIleArgGlnLul 683
QY 5940 AAGAATAAAGGATTTCAATAGAAAGAGAGAGAGTCAATATGCTCTGTGGACAGTG 5999
Db 683 ysaG1uIleLysG1yIleGlnLeuG1yLysG1uValLysLeuSerLeuPheAlaAspA 703
QY 6000 ACATGTTGTATATTAGAAACCCCATGCTCTCAGGCCAAAACCTCTTAACCTGATA 6059
Db 703 spHeIleValIyTrLeuGluAsnProIleValSerAlaGlnAsnLeuLeuLysLeuIleS 723
QY 6060 GCAACTTCAGCAAGTCTCAGACACAAAATCAATGTGCAAAAATCACAAGATTTCTAT 6119
Db 723 ersnPheSerLysValSerG1yTyrLysIleAsnValGlnLysSerGlnAlaPheLeuT 743
QY 6120 ACCGCAATTAATACAAACAGACAGCAAAATCATGAGTGAAGTCTCATTAACAATGCTA 6179
Db 743 yThrAsnAsnArgGlnThrG1uSerGlnIleMetG1yG1uLeuProPheThrIleAlas 763
QY 6180 CAAGAAGTAATAATCTGAGTAACAACTTACAGGAGCAGCTAGGAAGCTTCCAGG 6239
Db 763 erLysArgIleLysTyrLeuG1yIleGlnLeuThrArgAspValLysAspLeuPheLysG 783
QY 6240 AGAAGTACAAACCACTGATCAGAAATTAAGAGAGACACAAACAAATGAAAAATTC 6299
Db 783 LysnThrLysProLeuLeuLysG1uIleLysLysaspThrAsnLysTrpLysAsnIleL 803
QY 6300 CATGCTCAGATAGTAAGAATCAT-----GAAAATGCCATATCTGCCAAAGTAATT 6352
Db 803 rocYsSerTrpValG1yArgIleAsnIleValLysMetAlaIleLeuProLysValIleT 823
QY 6353 ATGATTCATGCTGATCCCATCAGATACCATTTGACTTCTTCACAGATTTGAAAAA 6412
Db 823 yArArgPheAsnAlaIleProIleLysLeuProMetThrPhePheThrG1uLeuGluLysT 843
QY 6413 CAACTTAATTAATTCATATGGAACCAAAAAAAGAGCCACAGACCAAGACAAATCTTAAGC 6472
Db 843 hrThrLeuLysPheIleThrPasnGln-LysArgAlaArgMetAlaLysSerIleLeuSer 862
QY 6473 AAAAAGAACAAAGCTGGAGGTATATCATCTGACTTAAACTATATCTAATAGGCTACA 6532
Db 863 GlnLysAsnLysAlaG1yG1yIleThrLeuProAspPheLysLeuTyrTyrLysAlaThr 882
QY 6533 GTAAACCAAACTGCATGCTATGATACCAAAACAGATATATACCAATGACAGACACA 6592
Db 883 ValThrLysThrAlaTrpTyrTyrGlnAsnArgspIleAspGlnTrpHisArgThr 902
QY 6593 GAGACCTCAGAAATTTAC-ACCTGCAATCTACATCATCTGATCTTGTGCAAACTGACAAA 6651
Db 903 GluProSerGluIleMetProHisIleTyrAsnTyrLeuIlePheAspLysProGluLys 922
QY 6652 AACACGCAATGGAAGAAAGATTCCTTATTAATATGTTGTGAAAAAAGTGGCTAGCC 6711
Db 923 AsnLysGlnTrpG1yLysAspSerLeuPheAsnLysTrpLysTrpGluAsnTrpLeuAla 942
QY 6712 ATATGCGAAGAGCGAAACCTGCATCCCTCTTACACCTTATCAAAAGTATCTACAGA 6771
Db 943 IleCysArgLysLeuLysLeuAspProPheLeuThrProTyrThrLysIleAsnSerArg 962
QY 6772 TGAATTAAGACTTAATATATAGACATATAAACCATAAAAACCA-GAAGAAGAACTAGGC 6830
Db 963 TrpIleLysAspLeuAsnValLysProLysThrIleLysThrLeuGlnGluAsnLeuG1y 982
QY 6831 AATACATTTGAGATATGACATGGGCAAGACTTCATGACTTAAGAACCAAAAGCAATG 6890
Db 983 IleThrIleGlnAspIleG1yValG1yLysAspPheMetSerLysThrProLysAlaMet 1002
QY 6891 GCAACAAAACCCAAATAGACAGTGGGATCTGATTAATATAGACTCTGCACAGCA 6950
Db 1003 AlaThrLysAspLysIleAspLysTrpAspLeuIleLysLeuLysSerPheCysThrAla 1022
QY 6951 AAAAAAACTGTCATCAGAGTGAACAAGCAACTACAGATGGGAGAAAAATTTTGGCAAT 7010
Db 1023 -LysGluThrThrIleArgValAsnArgGlnProThrThrTrpGluLysIlePheAlaThr 1042
QY 7011 CTATCGATCTGACAAAGGCTAATATATCCAGAGATCTACGAAGACTTAACAAATTTACA 7070
Db 1042 rTyrSerSerAspLysG1yLeuIleSerArgIleTyrAsnGluLeuLysGlnIleTyrLys 1062
QY 7071 GAAAAA-----AACACCCCGTCAAAATATAGGGAAGATAGAGAGAGACACTTCTCAA 7126
Db 1062 sLysLysTrpAsnAsnProIleLysLysTrpValLysAspMetAsnArgHisPheSerLys 1082
QY 7127 AGAAGACATTTATGACACCAACAACAATATGAAAAAAACCCTCATCATATGGTGGTTAG 7186
Db 1082 sGluAspIleTyrAlaIleAlaLysLysHisMetLysLysCysSerSerSerLeuAlaIleAr 1102
QY 7187 AGAAATGCCAAAAAACAACGAGTGACATACCATCTCATCTGATTTGAATGGTGATCAC 7246
Db 1102 gGluMetGlnIleLysThrThrMetArgTyrHisIleThrThProValArgMetAlaIleI 1122
QY 7247 TAAAGGTGAGAAACAAACAATGTGAGAGATGAGGAGAAATGGAACACTTTTCCA 7306
Db 1122 eLysLysSerG1yAsnAsnArgCysTrpArgG1yCysG1yGluIleG1yThrLeuLeuH 1142
QY 7307 CTGTGGTGGGAATGTAAATTAAGTCAACCATTTGTGGAAGACAGTGTGAGATTCCTTAA 7366
```



```

Db      1142 scystriprrpascyslysluvaljclnproleutrrpsserValtrpgrhneleuar 1162
QY      7367 GGAATACAGAACCAAAATATCATTTGACCAGCAATCCATTACTAGATATATACCAAA 7426
Db      1162 gaspleugluleuglulilepropheasproalalleproleuleuglylleyrrproly 1182
QY      7427 GGAATTAATATCTTTTATTATTAAGACACATCCACACATATGTTTATTGACACTGAT 7486
Db      1182 sasprtyrlyssercyscyslyrlyssaprhrcysttrhrcgmelPheillealaleuPh 1202
QY      7487 CCAATAGCAAGACACTGGACCAACCAATGTGCATGATGATGACTGGATAGAA 7546
Db      1202 ethrillealalystrhtrpshnclnprolyscsprotinmetiliasprtrilleysly 1222
QY      7547 AACATGCGACATATACCCATGAAATCTATGACGACCAATAAAAG-GATGAGTTCAATGC 7605
Db      1222 smettrphtstletytrthmetcglutyrtrhlaalallelyssasnapcluphellese 1242
QY      7606 CTTTGACAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7665
Db      1242 rhvevalgllytrhtrpmetlyslengluthrillelleuleuserlyleuserglengl 1262
QY      7666 GAAACCAACACACACATCTCTCACTTGTGAAGTGGAGT 7705
Db      1262 nlystrhlyshlsarglilepheserleulleglylyasn 1275

```

RESULT 11

012881 PRELIMINARY; PRT: 1275 AA.

```

AC      012881;
DT      01-NOV-1996 (TREMBLrel. 01, Created)
DT      01-NOV-1996 (TREMBLrel. 01, last sequence update)
DE      01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE      Retrotransposable Ll element LRE2 from chromosome 1q.
OS      Homo sapiens (Human)
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX      NCBI_Taxid=9606;
RN      (1)
RP      SEQUENCE FROM N.A.
RC      TRANSPOSON-LRE2;
RX      MEDLINE=95004577; PubMed=7920631;
RA      Holmes S.E., Dombroski B.A., Krebs C.M., Boehm C.D., Kazanian H.H.;
RT      "A new retrotransposable human Ll element from the LRE2 locus on
RT      chromosome 1q produces a chimaeric insertion.";
RL      Nat. Genet. 7:143-148(1994).
DR      EMBL: 009116; AAB60345.1;
DR      InterPro: IPR005135; Exo_endo_phos.
DR      InterPro: IPR000477; RVTse.
DR      Pfam: PF03372; Exo_endo_phos; 1.
DR      Pfam: PF00078; rvt; 1.
KW      RNA-directed DNA polymerase.
FT      VARIANT 1182 1182 E -> K.
SQ      SEQUENCE 1275 AA; 149064 MW; CEF4429261FE223 CRC64;

```

Alignment Scores:

```

Pred. No.: 0 Length: 1275
Score: 5169.00 Matches: 1054
Percent Similarity: 78.908 Conservative: 79
Best Local Similarity: 73.408 Mismatches: 137
Query Match: 9.888 Indels: 173
Gaps: 7

```

US-10-083-853b-2 (1-29921) x 012881 (1-1275)

```

QY      3428 ATGCGAGATCAATTCACACATTAATTAATTAATTAATTAATTAATTAATTC 3487
Db      1 MetrhnglySerAsnSerHisIleThrIleuThrIleuAsnIleAsnIleuAsnSer 20
QY      3488 CCAATTAAGACACAGATGCGCAATGATTAAGACAGACACACACACACACACACACAT 3547
Db      21 AlallelysalghlsarglenuAlasertrilleysserglinspserovalcyscys 40

```

```

QY      3548 ATTACAGAGGCCCATCTCACATGAAGACACACATAGCTCAAAATTAAGGATGAGG 3607
Db      41 lilegluluthrhisleumetcysargasprthhisargleuylsileysgllytrparg 60
QY      3608 AAGATTACCAGTAATTAAGAAACAAACAAAAAAGCGGGGTCCATTCCTAGCTCTG 3667
Db      61 lylsiletyrclnlnlansnglyslgn--Lysylsalgllyalalalleuvalser 79
QY      3668 GATTAACAGACTTTTAACCAACAAATGATCAAAAGAGACAAAGAGAGCCATTACATATG 3727
Db      80 Asprlyththasprheylsprottrhlysllelysarasprlyslucllyhstlytlemet 99
QY      3728 GTAAGCATCAATGAGACAGAGAGCTACTATCTTAATTAATTAATTAATTAATTAATTA 3787
Db      100 Vallylgllyserllelenglncgluleuthrilleleuasnillelyrhlaprosanthr 119
QY      3788 GAGACACCAAGATTCATTAAGACAGAGTCTTAAGAGCTTCAAAAGACACTTGCCTCCAC 3847
Db      120 GlyAlaprotargpheillelysglnvalleuserasprleuglntrgasprleuspserhis 139
QY      3848 ACAATATAGTGGAGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3907
Db      140 Thrleullemetly-----Asprheasnthrproleuserlleleu 153
QY      3908 GCGACATCAATGACAGAGAAATTAACAGATATCCAGAGTGAAGTGAAGTCTGGAC 3967
Db      154 Aspargsertharglinsvalasnlyasprthglngluleuasnseralaleuhis 173
QY      3968 CAAGCGACCTAATATATATCTACAGAACTCCCAACCCCAATCAACAGATATACAGTC 4027
Db      174 GlnAlasprleulleasprlleyrargtrhtrhlnhisprolysserthrluyltrhphe 193.
QY      4028 TTCTCGACATCAATTAACCTATTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4087
Db      194 PheserAlaprohishtshtrlyrserlyslleasprhisillevalglyserlyslaleu 213
QY      4088 CTCACGAATGCAAAAGAACAGAAATCTTAACAAACAGTCTCTACAGTACAGTGAATC 4147
Db      214 Leuserlyscylslyrgrhrcglullelthrasntlyleuserasphisseralalle 233
QY      4148 TATTTAAGCATGAGATTAAGAAACACACTCAAAACACAACTCATGGAAGCAGAAC 4207
Db      234 Lylslengluleuarglillelyssasnleuthrglnserargserthtrhtrpysleuasn 253
QY      4208 AACCTGCTCTGAATGACTACTGCGTAATTAACAAATGAAGGCAAAATTAAGATGTC 4267
Db      254 Asnleulleuasnasprrtyrtrpvalhlsasnnglmetlysalglulileysmetrpe 273
QY      4268 TTTGAACCAATGACAAACAGACACATGACCAATCTCTGGGCGATATTTAAGCA 4327
Db      274 PhegluthrasnglunasnlyssasprthtrtyrglnasnleuthrpsalaphelysAla 293
QY      4328 GTCTGAGAGGGAATTTATGACACTAGATGGCTCAAGAGAAAGAGAGGAATTAATTA 4387
Db      294 ValCysargglylyshellealleuasnAlatylLysarglysluclnluargserlysls 313
QY      4388 ATAGACACCTTAACATCAATTAAGAAAGAAAGAAAGAAAGAGCAAAATTAATTAAGAA 4447
Db      314 lileasprthleuthserclnleuysgluleuglulysglnlglulhthrserysls 333
QY      4448 GCTAGCAGAGACAAAGAAATTAAGATCAGACGAGAACTGAGAGAGATTAAGACACAA 4507
Db      334 AlaserargarglncglulthrlyslleargAlagluLeuLyglulleglulthrcln 353
QY      4508 AAGCCCTTCAATTAATCAATGATCCAGAGCTGCTTTTGAAGAAAGATACGCAAAAT 4567
Db      354 LysThrleuglnlys--lleasnglusertrrgsertrprphepgeglnuaglleasnlysl 373
QY      4568 ----AGACACATAGACAGACTAATTAAGAGAAAGAGAGAGAGATCAAGAGATGCAAT 4623
Db      373 easprgrproleuAlatargleullelyslslysarnglulysasnnglnlleasprthrll 393

```


QY 4624 AAAAAATGATTAAGGGATATTCACACCGATCCACAGAAATACAACTATATTCAGAGA 4683
|||||
Db 393 eulysasnAsplysGlyAspIlethrhrasprothrInglIleGlnThrThrIleatrgL 413
QY 4684 ATATTATAACACCTCTATGCAATTAACAGAAATCTAGAAGAAATGATTAATTCCT 4743
|||||
Db 413 uTyrrLysShIsleuTyrrLAsnLysLeuGlnAsnLeuGlnGluMetAspThrPhele 433
QY 4744 GGACACATATATGTCCTGATATGACCTGGGGGACAGAACAAAAGGGGATGATGCAGAA 4803
|||||
Db 433 unsp----- 434
QY 4804 ATAAAAAGACAAAGACAAAGATATGTTGGAAGTAGGGGTACAGGGGCAACTTGCCTT 4863
434 ----- 434
QY 4864 AATGCAGAAAGGCCCTGAGCTTTACACCCCTCTGTATTATTATTAAGCAAAAGATAGC 4923
434 ----- 434
QY 4924 GAGAGGTGAGTTGGAAGAAGAGGTGACGTGTAGGTCCAGAGTAGGCTGCAGAGACTGC 4983
434 ----- 434
QY 4984 ATTCTCAAAAGATAGCTCTAGATGTCCAGTAGATTAACCTCAGAGGCCAGTGCAGG 5043
434 ----- 434
QY 5044 GAGTGAATGGCCCTCAGCAAACTTCTTAGGCGACAGGACAGAAATAGTTTCCCACTTCT 5103
434 ----- 434
QY 5104 GATATCAGATAAACAATTGTTGTTGATCAAGTAGGCTCCAGTGGAAATGCTGAGTTGG 5163
434 ----- 434
QY 5164 TCATGATCCCTTTGGCCTTTTGGCTCCCAAAACATACACCTCTCAAGACTAAACCA 5223
435 ----- 435
QY 5224 GGAAGAATCAATCCCTGATATACAGTAACAGATTTCTAAATGAGACAGTAATGA 5283
|||||
Db 443 nglugluValGluSerLeuAsnGlyProIlethrLysSerGlnIleValAlaIleIleAs 463
QY 5284 TACCTACCAACCAAAAAAGATCCAGAGCAGAGGATTCACAGCCAAATTTACACAGAG 5343
463 nserIleuproThrLysLysSerProGlyProAspGlyPheThrAlaGluPheTyrrGlnat 483
QY 5344 GTACAAAGAGAGCTGTACTATCTTCTGAAACTATTCCAAAAAATAGAA--AATGG 5400
483 gTyrrLysGlnGluLeuValProPheIleuLeuLysLeuPheGlnSerIleGlnLysGlu- 503
QY 5401 GAATCTCTCTACTATTTTACGAGGCCGACATCATCTCGATACCAAAACCTAGCGATG 5460
|||||
Db 503 LylleuProAsnSerPheTyrrLysLysSerIleIleuLeuProLysProGlyArg 523
QY 5461 ACACACAAAGAGAGAAATTTACAGGCCCATATCCCTGATGACATGATGATGATAATTC 5520
523 spThrThrLysLysGlnAsnPheArgProIleSerLeuMetAsnIleAspAlaLysIleL 543
QY 5521 TCATAAATAATCTGGCAAAACCAATCCAGAGCAGATCAAAAGCTTATCTACCATATC 5580
543 eulsnLysIleLeuAlaAsnArgIleGlnIleHisIleLysLysIleHisIleAspG 563
QY 5581 AATTTGGCGTCATCCCTGGGATGCAGAGCTGTTCAAATATATGCAATCAATTAATAG 5640
563 lValAlaGlyPheIleProGlyMetGlnGlyTrrPheAsnIleArgLysSerIleAsnValI 583
QY 5641 GCATCATTAACAGAGAACCAATGACAAAACCAAGATATATCTCAATGATGACAGAAA 5700
583 lGlnIleHisIleAsnArgAlaLysAspLysAsnHisSerIleIleSerIleAspAlaLul 603
QY 5701 AGGCTTTGTCAAAATTCACAGCCCTTCATGCTAAAAATTTCTAGTAACCTAGGTATCG 5760

Db 603 ysaIlePheAspLysIleGlnGlnProPheMetLeuLysThrIleAsnLysLeuGlyIleA 623
QY 5761 ATGGAATGTATTCACAAATATAAGACTTTTATAC-AAACCCACAGCCCAATATCTATC 5819
|||||
Db 623 spGlyThrTyrrPheLysIleIleArgAlaIleIleTyrrAspLysProThrAlaAsnIleIle 643
QY 5820 TGAATGGGCAAAACCTGAGCAATCCCTTGAACACTGGCACAGCAAGATGGCCCTC 5879
643 eulsnGlnGlnLysLeuGlnAlaPheProLeuLysThrGlyThrArgGlnGlySerProL 663
QY 5880 TCTCACACCTCTATTCACAGATCTATTTGGAAGTTCTGGCCAGGCAATCAGCAATAGA 5939
663 euserProLeuLeuPheAsnIleValLeuGlnValLeuAlaIleArgGlnGlnLul 683
QY 5940 AAGCAATTAAGGCTATTCAAATAGAAAGAGAGAGATCATATTTCTCTGTTGCAGATG 5999
683 yscIuIleLysGlyIleGlnLeuGlyLysGlnGluValLysLeuSerLeuPheAlaAsp 703
QY 6000 ACATGTTTGTATATTTTGAAGAACCCCATGCTCAGGCCCAAAACCTTAAAGCTGATAA 6059
703 spMetIleValIlyrLeuGlnAsnProIleValSerAlaGlnAsnLeuLysLeuIles 723
QY 6060 GCAACTTCAGCAAAAGTCTCAGAGACACAAATCAATGTGCAAAATCACAAGACTTCTAT 6119
723 eAsnPheSerLysValSerGlyTyrrLysIleAsnValGlnLysSerGlnAlaPheLeu 743
QY 6120 AGCCGATATATAGACAAACAGACAGACCAATTCATGAGTGAACITCTATTCACAAITTCCTA 6179
743 yrrhrAsnAsnArgGlnThrGlnSerGlnIleMetGlyGlnLeuProPheThrIleAlas 763
QY 6180 CAAAGAGAAATAAATACCTAGATACACTTACAAAGGACAGCTAGCACTCTCAAG 6239
763 eTyrrAspGlyIleLysTyrrLeuGlyIleGlnLeuThrArgAspValLysAspLeuPheLysG 783
QY 6240 AGAAGCTACAAACACCTGATCAGAGAAATAAGAGAGACACAAACAAATGAAAAATTC 6299
783 lAsnTyrrLysProLeuLeuLysGlnIleLysGlnAspThrAsnLysTyrrLysAsnIleP 803
QY 6300 CATGCTCAGATATAGATATCAT-----GAAATGGCATTCCTCCCAAGTAATT 6352
803 rocYsSerTrpValGlyArgIleAsnIleValLysMetAlaIleLeuProLysValIle 823
QY 6353 ATAGATTCAGTGTGATCCCATCAGCTACATGATGATCTTCTCAGAAATGGAATAA 6412
823 yrrArgPheAsnAlaIleProIleLysLeuProMetThrPhePheThrGlnLeuGlnLysT 843
QY 6413 CAATTTAAATTCATATGGAACCAAAAAAGAGCCACAGAGCAAGCAAGCAATCTTAAGC 6472
843 hrThrLeuLysPheIleTrpAsnGln-LysArgAlaArgIleAlaLysSerIleLeuSer 862
QY 6473 AAAAAGAACAAAGCTGAGGTATCATCTACCTGATCTTAACATATCTTAAGGCTACA 6532
863 GlnLysAsnLysAlaGlyGlyIleThrLeuProTyrrPheLysLeuTyrrLysAlaThr 882
QY 6533 GTAACCAAACTGCATGCTACTGTACCAAAACAGATATATAGCAATGGAAGAGACA 6592
883 ValThrLysThrAlaTrpTyrrTrpTyrrGlnAsnArgLysPheGlnIntrrPAsnArgThr 902
QY 6593 GAGACCTCAGAAATTC-ACTCAGATCTACATCCATGTGATCTTTGACAAACCTGACAAA 6651
903 GluProSerGlnIleMetProHisIleTyrrAsnTyrrLeuIlePheAspLysProGlnLys 922
QY 6652 AACACAGAAATGAAAAAGATTCCTCTTAAATAAGGCTTGGAANAACCTGGCTAGCC 6711
923 AsnLysGlnTrpGlyLysAspSerLeuPheAsnLysTrpCysTrpGlnAsnTrpLeuAla 942
QY 6712 ATATGACAGAAAGCTGAACTGATCCCTCTACCTATATACAAAGTAATCTACAGA 6771
943 lIleCysArgLysLeuLysLeuAspLeuPheLeuThrProTyrrThrLysIleAsnSerArg 962
QY 6772 TGAATTAAGACCTTAAATATAGACATAAAAACATAAAAAACA-GAAGAAAACCTAGGC 6830

Db 963 TrrpIleLysAspLeuAsnVallLysProLysThrIleLysThrLeuGluLysLeuGly 962
QY 6831 AATACCATTCAGATATGACATGGGCAAGACTTCATGACTTAACACCAAGCAATG 6890
Db 983 IletHrIleGlnAspIleGlyValGlyLysAspPheMetSerLysThrProLysAlaMet 1002
QY 6891 GCAACAAAGCCAAATAGACAGTGGATCTGATTAACTATAGACTTTCGACAGCA 6950
Db 1003 AlAtHrLysAspLysIleAspLysTrpAspLeuIleLysLeuLysSerPheCysThrAla 1022
QY 6951 AAAAAAAGCTGACAGTGAAGCAAGCAACCTGACAGATGGGAGAAAATTTTCGAA 7010
Db 1023 -LysGluThrHrIleArgValAsnArgGlnProThrHrIleGluLysIlePheAlaH 1042
QY 7011 CTATCGATCTGCACAAAGGCTAAATATCCAGATCTGACAGAACTTAACAAATTACAA 7070
Db 1042 rLysSerSerAspLysGlyLeuIleSerArgIleTyraGlnLysLeuGlnIleTyrl 1062
QY 7071 GAAAAA---AACACCCCGTCAAAATATGGCAAGATATGACAGCAACTTTCGAAA 7126
Db 1062 sLysLysThrAsnAsnProIleLysLysTrpAlaLysAspMetAsnArgHisPheSerL 1082
QY 7127 AGAAGCATTTATGCGCAACAAACATATGAAAAAACCCTGATTCATGGTGGTGG 7186
Db 1082 sGluAspIleTyrlAlaAlaLysLysHisMetLysLysCysSerSerLeuAlaIleAr 1102
QY 7187 AGAAATGCAAAACAAACACAGTACATACCATTCATGCTAGTTAGAATGGTGATCAC 7246
Db 1102 gGluMetGlnIleLysThrHrMetArgTyrlHisLeuThrProValArgMetAlaIleAr 1122
QY 7247 TAAAAAGTACGAAACACAAATGCTGGAGAGATGTGAGAAATATGAAACCTTTTCCA 7306
Db 1122 eLysLysSerGlyAsnAsnArgCysTrpArgGlyGlnIleGlyThrLeuLeuHr 1142
QY 7307 CTGTGGTGGAGATGAAATTAATGATTCACATTCATGTCGAGACAGCTGGAGATCTCTTA 7366
Db 1142 sCysTrpTrpAspCysLysLeuValGlnProLeuThrLysSerValTrpArgPheLeuAr 1162
QY 7367 GGATCTAGAACCAAGAAATATCATTTGACCCAGCAATCCCATTAAGTATATACCCAAA 7426
Db 1162 gAspLeuGlnLeuGlnIleProPheAspProAlaIleProLeuLeuGlyIleTyrlProG 1182
QY 7427 GGAATATTAATCATCTATTATAAGACATGACACATATGCTTATATGACGACCTGAT 7486
Db 1182 uAspLysLysSerCysCysTyrlLysAspThrArgMetPheIleAlaLeuHr 1202
QY 7487 CACAATAGCAAAAGACTTGGACCAACCCAAATGTCATCATGATAGACTGATTAAGAA 7546
Db 1202 eHrIleAlaLysThrTrpAsnGlnProLysCysProThrMetIleAspTrpIleLysL 1222
QY 7547 AACATGGCATAATACCATGAAATATCTATGACAGCCATTAAG- GATGAGTTGATGTC 7605
Db 1222 sMetTrpHisIleTyrlThMetGluTyrlTyrlAlaIleLysAsnAspGlnPheIleSe 1242
QY 7606 CTTTGACAGATATGAGTGAAGCTGGAACCATTCATTCACAAACTACACAGAAACA 7665
Db 1242 rPheValGlyThrTrpMetLysLeuGlnHrIleIleLeuSerLysLeuSerGlnLeuG 1262
QY 7666 GAAAAACAAACACACATGTTCTCACTTGTAAAGTGGAGAT 7705
Db 1262 nLysThrLysHisArgIlePheSerLeuIleGlyLysAsn 1275

RESULT 12
000378 PRELIMINARY; PRT; 1275 AA.
AC 000378;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative p150.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97285120; PubMed=9140393;
RA Sassaman D.M., Dombroski B.A., Moran J.V., Kimberland M.L., Neas T.P.,
RA DeBerardinis R.J., Gabriel A., Swergold G.D., Kazazian H.H. Jr.;
RT "Many human LI elements are capable of retrotransposition.";
RL Nat. Genet. 16:37-43(1997).
DR EMBL; U93574; AAC51279.1; -
DR InterPro: IPR005135; Exo_endo_phos.
DR InterPro: IPR000477; RVTse.
DR Pfam: PF003372; Exo_endo_phos; 1.
DR Pfam: PF00078; rvc; 1.
DR RNA-directed DNA polymerase.
SQ SEQUENCE 1275 AA; 148879 MW; F006F74BBB872B87 CRC64;
Alignment Scores:
Pred. No.: 0 Length: 1275
Score: 5168.00 Matches: 1053
Percent Similarity: 78.69% Conservative: 77
Best Local Similarity: 73.33% Mismatches: 140
Query Match: 9.88% Indels: 173
DB: Gaps: 7
US-10-083-853B-2 (1-29921) x 000378 (1-1275)
QY 3428 ATGCGAGATCAAAATTCACACATATATTAATTAACCTTAATGTAATGGCTAAATTC 3487
Db 1 MetThrGlySerAsnSerHisIleThrIleLeuThrLeuAsnIleAsnGlyLeuAsnSer 20
QY 3488 CCAATTAAGACACACAGACTGGCAAAATGGATTAAGAGTCAAGACCATCAGTGCCT 3547
Db 21 AlaIleLysArgHisArgLeuLysSerTrpIleLysSerGlnAspProSerValLysCys 40
QY 3548 ATTCGGAGGCCCATCTGCACAGAAAGACACACATATGCTGCTGCTGCTGCTGCTGCTG 3607
Db 41 IleGlnGluThrHisLeuThrCysArgAspTrpHisArgLeuLysIleLysGlyTrpArg 60
QY 3608 AAGATTTTACCAAGTAATGGAACCAAAAAAAGAGAGGGGTTGCAATCTAGTCTCT 3667
Db 61 LysIleTyrlGlnAlaAsnGlnLysGln--LysLysAlaGlyValAlaIleLeuValSer 79
QY 3668 GATTAACACGACTTTTAACCAACAAAGATCAAAAGACAGCAAGAGCCATTAACATG 3727
Db 80 AspLysThrAspPheLysProThrLysIleLysArgAspLysGlnIleLysTrpIleMet 99
QY 3728 GTAAAGGACATCATGAGCAAGCAAGAGCTAATCTTAATATGATCATGACACCAATACA 3787
Db 100 ValLysGlySerIleGlnGlnIleGluLeuThrIleLeuAsnIleTyrlArProAsnThr 119
QY 3788 GGAGCACCCAGATTCATATAAGCAAGTCTTAAAGACTTACAAAGAGACTTTGACTGCCAC 3847
Db 120 GlnAlaProArgPheIleLysGlnValIleLeuSerAspLeuGlnArgAspLeuAspSerHis 139
QY 3848 ACAATTAATGTGGGAGCTCAATTAATTAATTAACACTTAACACCCCACTGCCAATATTA 3907
Db 140 ThrLeuIleMetGly-----AspPheAsnThrProLeuSerThrLeu 153
QY 3908 GCGAGATCATAGTACAGCAAGAAATTAACAGATATATCAGAGATTTGAAGCTCTGAGC 3967
Db 154 AspArgSerThrArgGlnLysValAsnLysAspTrpGlnIleLeuAsnSerAlaLeuHis 173
QY 3968 CAAGCGAGCTAATAGATATCTACAGAACTCCSCACCCCAATTAACAGAGATATACACTC 4027
Db 174 GlnAlaAspLeuIleAspIleTyrlArgTrpHisLeuHisProLysSerThrGlnIleTyrlPhe 193
QY 4028 TTCTCAGATCATCAATTAACACCATTTTAATTAATGACAGATTAATTAAGTAAACACATC 4087
Db 194 PheSerAlaProHisHisThrTyrlSerLysIleAspHisIleValGlySerLysAlaLeu 213
QY 4088 CTCGCAAAATGCAAGAAAGAAAGAAATCTTAACAAACAGCTCTCAGACTACAGTCAATC 4147
Db 214 LeuSerLysCysLysArgThrGluIleIleThrAsnTyrlLeuSerAspHisSerAlaIle 233


```

Db      803  rocySerTrpValGlyArgIleasnIleValIysMetAlaIleLeuProIysValIleT 823
QY      6535  ATACATTCAGTGTACCCCATCAAGTACATGACTTCTTCCAGCAATTTGCAAAAA 6412
Db      823  yTAhgpheasnAlIleProIleIysLeuProMetThrPhePheThcIleuEngIlyst 843
QY      6413  CAACTTAATTTGATATGAAACCAAAAAAGAGCCACAGCCCAAGCAATCTTAAGC 6472
Db      843  hThrlleuIysPheIleTrpAsnGln-LysArgAlaIleAlaIleIysAlaIleLeuSer 862
QY      6473  AAAAAGAACAAAGCTGGAGGTATGATGCTACCTGACTTAACATATACATTAAGGCTTCA 6532
Db      863  GlnIysAsnIysAlaGlyGlyIleThrlleuProAspPheIysLeuTyTylAlaIleThr 882
QY      6533  GTAAACCAAACTGATGCTAGTACCAAAAAGATATATAGACCAATGAGCAAGAACAA 6592
Db      883  ValThrlYsthrAlaIleTrpIlyTrpIlyGlnAsnAspIleAspGlnTrpAsnArgThr 902
QY      6593  GAGACCTCAGAAATTAAC-ACAGCAATCTACATCCATGATCTTTGACAAACCTGACAAA 6651
Db      903  GluProSerGluIleMetProHisIleTyAsnTyIleuIlePheAspIysProGlnIlys 922
QY      6652  AACAGCAATGGAAAAAGATTCCTTATTTAATAGGTGTGAAAAACTGGCTAGCC 6711
Db      923  AsnIysGlnTrpGlyIysAspSerLeuPheAsnIysTrpCysTrpGluAsnTrpLeuAla 942
QY      6712  ATATGCAAAAAGCTGAAAGTGGATCCCTTCCCTACACTTATTCAAAAGTAACTCAGA 6771
Db      943  IleGysArgIysLeuIysLeuAspProPheLeuThrProTyThrIlyIleAsnSerArg 962
QY      6772  TGAATTAAGACTTAATATATAGACATTAACCAATAAACCCCA-GAAGAAACCTTAGGC 6830
Db      963  TrpIleIysAspLeuAsnValArgProIlyThrIleIysThrLeuGlnGluAsnLeuGly 982
QY      6831  AATACCATTCAGATATGAGCATGGCAAGACTTCATGATTAACCAACCAAGCATG 6890
Db      983  IleThrlleGlnAspIleGlyValGlyIysAspPheMetSerIysThrProIysAlaMet 1002
QY      6891  GCACAACAAAGCCAAATATAGCAAGTGGATGATGATTAACATATGACTTTCGACAGCA 6950
Db      1003  AleThrlYsAlaIysIleAspIlystrpAspLeuIleIysLeuIysSerPheCysThrAla 1022
QY      6951  AAAAAAACCTGCTCAGAGTGAACAAGCAACCTACAGATGGAGAAAAATTTTGCAT 7010
Db      1023  -LysGluThrThrIleAlaArgAlaAsnArgGlnProThrThrIlyIysIlePheAlaThr 1042
QY      7011  CTATCGATCTGACAAAGGCTAATATATCCAGATCTAGAGAACTTAACAAATTTTCAA 7070
Db      1042  rTyIserSerAspIysGlyLeuIleSerArgIleTyAsnGlnIleuIysGlnIleTyIly 1062
QY      7071  GAAAAA-----ACAAACCCCTCAAAAATATGGGCAAGAGATATGAGCAACACTTCAAA 7126
Db      1062  sIysIysThrAsnAsnProIleIysIysTrpAlaIysAspMetAsnArgHisPheSerIly 1082
QY      7127  AGAAGACATTTATGACACCAACAACATATGAAAAAACTCATCATCATTTGTCGTTAG 7186
Db      1082  sGluAspIleTyAlaIleAlaIysIysHisMetIysIysCysSerSerIleuAlaIleArg 1102
QY      7187  AGAAATGCAAAACCAACACAGTACATACCATTCATGCTAGTATGAAATGGATGACAC 7246
Db      1102  gGluMetGlnIleIysThrThrMetArgCysHisPheThrProValArgMetAlaIleIle 1122
QY      7247  TAAAAAGTCAGAAACAAATGCTGAGAGATGTGAGAAATAGCAACACTTTCCTCA 7306
Db      1122  eIysIysSerGlyAsnAsnArgCysTrpArgGlyCysGlyGluIleIlyThrLeuLeuHis 1142
QY      7307  CTGTGGTGGCAATGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 7366
Db      1142  scYsTrpTrpAspCysIysLeuValGlnProLeuTrpIlySerValTrpArgPheLeuArg 1162
QY      7367  GGAATCTGAACCAAGAAATATATGACCAAGCAATCCCATTAAGTATATATCCAAA 7426

```

```

Db      1162  gasPLeuGluLeuGluIleProPheAspProAlaIleProLeuLeuGlyIleTyProIly 1182
QY      7427  GGAATTAATTAATCTTATTAATTAAGACATGACATGATGTTATTAAGACACTGAT 7486
Db      1182  sAspTyIlySerCysCysTyTrpIysAspThrCysThrArgMetPheIleAlaIleAlaLeuPh 1202
QY      7487  CACAATAGCAAAAGCTTGGAAACCAACCAAAATGTCCATCACTGATAGACTGATTAAGAA 7546
Db      1202  eThrlleAlaIysThrIleTrpAsnGlnProIysCysProThrMetIleAspTrpIleIysIly 1222
QY      7547  AACATGCAATATATACACATGAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7605
Db      1222  sMetTrpHisIleTyThrIleMetGluTyTrpAlaAlaIleIysAsnAspGluPheMetIse 1242
QY      7606  CTTTGCAAGATATGATGTAAGAGCGAAGCAACCATCATCTGCAACCAATACCAAGAACAA 7665
Db      1242  rPheValGlyThrTrpMetIysLeuGluThrIleIleLeuSerIysLeuSerGlnGluI 1262
QY      7666  GAAACCAAAACCAACCATGTTCTGACCTTGAATGAGGAGT 7705
Db      1262  nIysThrIysHisArgIlePheSerLeuIleGlyIysn 1275

RESULT 13
ID      000368  PRELIMINARY;  PRT: 1275 AA.
AC      000368;
DT      01-JUL-1997 (TReMBLrel. 04, Created)
DR      01-JUL-1997 (TReMBLrel. 04, Last sequence update)
DE      01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE      Putative p150.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OX      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_Taxid:9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE-97285120; PubMed-9140393;
RA      Sasseman D.M., Dombroski B.A., Moran J.V., Kimberland M.L., Naas T.P.,
RT      Debernardinis R.J., Gabriel A., Swergold G.D., Kazazian H.H., Jr.;
RL      "Many human LI elements are capable of retrotransposition.";
DR      EMBL: U93568; AAC51269.1; -.
DR      InterPro: IPR005135; Exo_endo_phos.
DR      InterPro: IPR00477; RVTse.
DR      Pfam: PF03372; Exo_endo_phos; 1.
DR      Pfam: PF00078; Ixt_1.
KW      RNA-directed DNA polymerase.
SQ      SEQUENCE 1275 AA; 149078 MW; 3BBC3B2DC2E06B61 CRC64;

Alignment Scores:
Pred. No.: 0 Length: 1275
Score: 5163.00 Matches: 1053
Percent Similarity: 78.83% Conservative: 79
Best local Similarity: 73.33% Mismatches: 138
Query Match: 9.87% Indels: 173
Gaps: 7

US-10-083-853b-2 (1-29921) x 000368 (1-1275)
QY      3428  ATGGCAGACCAAAATTCACCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTC 3487
Db      1  MetThrIysSerAsnSerHisIleThrIleLeuThrLeuAsnIleAsnGlyLeuAsnSer 20
QY      3488  CCAATTAAGACACACAGACTGCAAAATTTGATTAAGAGTCAAGACCATGCTGCTGT 3547
Db      21  AlalIeIysArgHisArgIleAlaIleAspTrpIleIysSerGlnAspProSerValCysCys 40
QY      3548  ATTTCAGAGGCCCATCTCACATGAAAAAGACACACATAGGCTCAAAATTAAGAGATGGAG 3607
Db      41  IleGlnIleuThrHisLeuThrIysArgAspThrHisArgIleuIysIleIysGlyTTrpArg 60
QY      3608  AAGATTTCACCAAGTAATGAAAAACAAAAAAGCAGGGGTGCAATCTAGTCTCT 3667

```

Db	61	LysIleTyrGlnAlaAsnGlyLysGln---	LysLysAlaGlyValAlaIleLeuValSer	79
Qy	3668	GATAAAACAGACTTTAAACCAACAAAGATCAAAAGACAGACAAAGAGCCATTACATAATG		3727
Db	80	AspLysThrAspPheLysProThrLysIleLysArgAspLysGluGlyHisTyrIleMet		99
Qy	3728	GTAAGGCATCAATGGAAACAGAGAGCTAACTATCCTAAATATATACATGACCCCAATACA		3787
Db	100	ValLysGlySerIleGlnGlnGluLeuThrIleLeuAsnIleTyrAlaProAsnThr		119
Qy	3788	GGAGCACCCAGATTCAATAAGCAAGTTCTTAGAGACCTCAAAAGAGACTTTGGCTCCAC		3847
Db	120	GlyAlaProArgPheIleLysGlnValLeuSerAspLeuGlnArgAspLeuAspSerHis		139
Qy	3848	ACAATAATAGTGGGAGTCTAAATAAATAGACACTTTAAACACCCCACTGCCAATATTA		3907
Db	140	ThrLeuIleMetGly-----	AspPheAsnThrProLeuSerThrLeu	153
Qy	3908	GGCAGATCAATGAGACAGAAAATTAAACAGGATATCCAGAGTTGAACTGAGCTCTGGAC		3967
Db	154	AspArgSerThrArgGlnLysValAsnLysAspThrGlnGluLeuAsnSerAlaLeuHis		173
Qy	3968	CAAGCGACCTAATAGATCTACAGAACTCCCAACCCCAATCAACAGAAATATACACTC		4027
Db	174	GlnAlaAspLeuIleAspIleTyrArgThrLeuHisProLysSerThrGluTyrThrPhe		193
Qy	4028	TTCTCAGCATCACATTACACTTATTTAAATTTGACCATGTAATTTTAACTAAACACTC		4087
Db	194	PheSerAlaProHisHisThrTyrSerLysIleAspHisIleLeuGlySerLysAlaLeu		213
Qy	4088	CTCAGCAATGCAAAAGACAGAAATCTCAACAAACAGTCTCTCAGACTACAGTGCATC		4147
Db	214	LeuSerLysCysLysArgThrGlnIleThrAsnTyrLeuSerAspHisSerAlaIle		233
Qy	4148	TATTTAGAACTCAGAAATTAGAAACTCACTCAAAATCACACACTACATGGAACCTAAC		4207
Db	234	LysLeuGluLeuArgIleLysAsnLeuThrGlnSerArgSerThrTrpLysLeuAsn		253
Qy	4208	AACTGCTCTGATGACTACTGGTAAATAAACAATAAGGCAAAATAAAGATGTC		4267
Db	254	AsnLeuLeuLeuAsnAspTyrTrpValHisAsnGluMetLysAlaAspIleLysMetPhe		273
Qy	4268	TTTGAACCAATGAGAAACAGACAAATGTACAGAACTCTCGGGGCATATTTAAAGCA		4327
Db	274	PheGluThrAsnGluAsnLysAspThrThrTyrGlnAsnLeuTrpAspAlaPheLysAla		293
Qy	4328	GTGTGTAGAGGAAATTTATAGACTAGATGCCTTACAAAGAGAAAGCAGGAATATCTAAA		4387
Db	294	ValCysArgGlyLysPheIleAlaLeuAsnAlaTyrLysArgLysGlnGluArgSerLys		313
Qy	4388	ATAGACACCTTAACATCACAAATTAAGAACTAGAGAAGAAAGACAAACAAATTCAAA		4447
Db	314	IleAspThrLeuThrSerGlnLeuLysGluLeuGlyGlnGlnGlnThrHisSerLys		333
Qy	4448	GCTAGCAGACAGCAAGAAATTAAGTACAGACAGAACTCAAGGAGATAGACACAA		4507
Db	334	AlaSerArgArgGlnGluIleThrLysIleArgAlaGluLeuLysGluIleGluThrGln		353
Qy	4508	AAAGCCCTCAATAATCAATCAATCCAGAGCTGGTTTTTCAAAAGATCAGCAAAAT		4567
Db	354	LysThrLeuGlnLys-IleAsnGluSerArgSerTrpPheGluArgIleAsnLysIle		373
Qy	4568	----AGACCACTAGACAGACTAATAAGAAAGAAAGAGAAAGTCAAGAGATGCAAT		4623
Db	373	eAspArgProLeuAlaArgLeuIleLysLysLysArgGluLysAsnGlnIleAspThrIle		393
Qy	4624	AAAAATGATAAGGGGATATCACCCGATCCCAAGAAATACAAACTATTATCAGAGA		4683
Db	393	eLysAsnAspLysGlyAspIleThrThrAspProThrGluIleGlnThrThrIleArgGln		413
Qy	4684	ATATTAAACACTCTCTCAATAAATAACTAGAAAATCTAGAGAAATGATAAATTCCT		4743
Db	413	uTyrTyrLysHisLeuTyrAlaAsnLysLeuGluAsnLeuGluMetAspThrPheLe		433
Qy	4744	GGACACATATCTAGCTGTATGGACCTTTGGGGGACAGACAAAGGGGGTGAATGCAGAA		4803
Db	433	uAsp-----		434
Qy	4804	ATAAAGACAAAGACAAAGAGATATGTTTGGAAAGTAGGGTTCAGGGGGCAACTTGCCTCT		4863
Db	434	-----		434
Qy	4864	AATGGACAAGGGCCCTGAGCTTTTACACCCCTCTGTATTTATTATGACAAAGAGATAGC		4923
Db	434	-----		434
Qy	4924	GAGAGGCTGATTGGAAAGAGAGGTCAGCTGTTAGTCCAGAGTAGGCTTGAAGACTGC		4983
Db	434	-----		434
Qy	4984	ATTCTCTCAACAATAGAGCTCTAGATGTCCAGTAGATAAACCCTCAAGGAGCCAGTGCAGG		5043
Db	434	-----		434
Qy	5044	GAGTGATGGCCCTCAGCAAACTTCTAGGGCAGGCACAGAAAGTAAGTTTGGCCACATCT		5103
Db	434	-----		434
Qy	5104	GTATTACGATATAACAGCTTTTGTGATCAAGTAGCTCCAGTCCAGTGAATGCTGAGTTGG		5163
Db	434	-----		434
Qy	5164	TCATGATCCCTTTGGCCCTTTTGGCTCCCAAAACACATACACCTCTCAGAGCTAAACCA		5223
Db	435	-----	ThrTyrThrLeuProArgLeuAsnGln	443
Qy	5224	GGAGAGACTCAATCCCTGAATATATACCAGTAACAGTCTCAAAATTCAGACAGTAATGA		5283
Db	443	nGluGluValGluSerLeuAsnArgIleThrGlySerGluIleValAlaIleIleAs		463
Qy	5284	TAGCTACCAACCAAAAAAGTCCAGGACCCAGCGATTCCACAGCCAAATTTCTACCAAG		5343
Db	463	nSerLeuProThrLysLysSerProGlyProAspGlyPheThrAlaGluPheTyrGlnArg		483
Qy	5344	GTACAAGAGAACTGGTACTATTCCTCTGAAACTATTCACAAAAATAGAAA---AATGG		5400
Db	483	gTyrMetGluGluValProPheLeuLeuLysLeuPheGlnSerIleGluLysGlu-G		503
Qy	5401	GAATCCCTCCCTAACTCACTTTTACGAGGCCAGCATCTCTGATACCAAAACCTAGCAGT		5460
Db	503	lyIleLeuProAsnSerPheTyrGluAlaSerIleIleLeuIleProLysProGlyArgA		523
Qy	5461	ACACAACAAAAGAGGAAATTTTCAGGCCCATATCCCTGTGATGAACATTTGATGTGAAAATCC		5520
Db	523	spThrThrLysLysGluAsnPheArgProIleSerLeuMetAsnIleAspAlaLysIleL		543
Qy	5521	TCATAAATACTGGCAACCAATCCAGCAGACATCAAAAAGCTTATCTTACCATCATC		5580
Db	543	euAsnLysIleLeuAlaAsnArgIleGlnGlnHisIleLysLysLeuIleHisaspG		563
Qy	5581	AAGTGGCGTCATCCCTGGGATGCAAGCTGGTTCAAAATATGCAATCAATAATCTAG		5640
Db	563	InValGlyPheIleProGlyMetGlnGlyTrpPheAsnIleArgLysSerIleAsnValI		583
Qy	5641	GCCATCACATAAAGACAGAAATTCACAAAAACCATGATTTCTCAATAGATGCAGAAA		5700
Db	583	IeGlnHisIleAsnArgThrLysAspLysAsnHisMetIleIleSerIleAspAlaGluL		603
Qy	5701	AGGCTTTGTCAAAATTCACAGCCCTTCATGCTAAATAATTCAGTAAGTACTAGGTATCG		5760
Db	603	ysAlaPheAspLysIleGlnGlnProPheMetLeuLysThrLeuAsnLysLeuGlyIleA		623
Qy	5761	ATGCAATGTATCTCAAAATATAGACTATTTATATAC-AAACCCACAGCAATATCATAC		5819
Db	623	spGlyThrTyrPheLysIleIleArgAlaIleTyrAspLysProThrAlaAsnIleIleL		643

```
QY 5820 TGAATGGGCAAACTGGAAGCATTCCTTTGAGAAGTGGCAAGAGGATGCCCTC 5879
Db 643 euAsnGlyGlnLysLeuGluAlaPheProLeuLysThrGlyThrArgGlnGlyCysProL 663
QY 5880 TCTCACCCTCTATTCAAGATACTATTGGAAGTCTGCCCGGCAATCAGGCATAGA 5939
Db 663 euSerProLeuLeuPheAsnIleValLeuGluValLeuAlaArgAlaIleArgGlnGluL 683
QY 5940 AAGAAATAAAGGGTATTCAAAAGAGAGAGGAAGTCATATTGCTCTGTTGGAGATG 5999
Db 683 ysgLulileLysGlyIleGlnLeuLysGluGluValLysLeuSerLeuPheSerAspA 703
QY 6000 ACATGTTTGTATATTAGAAACCCCATCGTCTCAGGCCCAAACTCCTTAAGCTGATAA 6059
Db 703 spMetIleValTyLeuGluAsnProIleValSerAlaGlnAsnLeuLeuLysLeuIleS 723
QY 6060 GCAACTTCAGCAAGCTCAGACACAAATCAATGTGCAAAANTCACAAGCATCTTAT 6119
Db 723 exAsnPheSerLysValSerGlyTyLeuIleAsnValGlnLysSerGlnAlaPheLeuT 743
QY 6120 AGCCCAATAATAGACAAACAGAGCCAAATCATGAGTGAAGTCTCATTCACAAATTCGTA 6179
Db 743 yrThrAsnAsnArgGlnThrGluSerGlnIleMetGlyGluLeuProPheThrIleAlaS 763
QY 6180 CAAAGAGATAAACTAGGAATACAACTTACAGGGACACGCTAGGAACCTCTCAAG 6239
Db 763 erLysArgIleLysTyLeuGlyIleGlnLeuThrArgAspValLysAspLeuPheLysG 783
QY 6240 AGAAGTACAAACCACTGATCAAGGAATAAGAGAGACACAAACAATGGAAACATTC 6299
Db 783 luAsnTyLeuProLeuLeuLysGluIleLysGluGluThrAsnLysTyLeuIlePhe 803
QY 6300 CATGCTCAGACATAGTAAGAATCAT-----GAAATGCCATCTGCCCCAAAGTAAAT 6352
Db 803 roCysSerTrpValGlyArgIleAsnIleValLysMetAlaIleLeuProLysValIleT 823
QY 6353 ATAGATTACGTGCTACCCCATCAAGCTACCATGCTTACAGTCTTCTCAGAAATTTGGA 6412
Db 823 yrArgPheAsnAlaIleProIleLysLeuProMetThrPhePheThrGluLeuGluLysT 843
QY 6413 CAACTTTAAATTCATATGGAACCAAAAGAGCCACAGACCAAGCAACATCTTAAGC 6472
Db 843 hrThrLeuLysPheIleTrpAsnGln-LysArgAlaArgIleAlaLysSerIleLeuSer 862
QY 6473 AAAAGACAAAGCTGGAGGTATCATGCTTACCTTAAACTACTATATAAGGCTACA 6532
Db 863 GlnLysAsnLysAlaGlyIleThrLeuProAspPheLysLeuTyTyLysAlaThr 882
QY 6533 GTAACCAAACTGCATGGTACTGGTACCAAAACAGATATATAGACCAATGGGAACAGA 6592
Db 883 ValThrLysThrAlaCysTyTrpTyGlnAsnArgAspIleAspGlnTrpAsnArgThr 902
QY 6593 GAGACTCAGAATATACACT-GCAATCTACATCCATCTGATCTTTGACAAACCTGCACAA 6651
Db 903 GluProSerGluIleThrProHisIleTyAsnTyLeuIlePheAspLysProGluLys 922
QY 6652 AACAAGCAATGAAAAGGATTCCTTATTATAAATGGTGTGGAAAACCTGGCTAGCC 6711
Db 923 AsnLysGlnTrpGlyLysAspSerLeuPheAsnLysTyTrpCysTrpGluAsnTrpLeuAla 942
QY 6712 ATATGAGAAAGCTGAAAGTGCATCCCTTACCTTACACCTTATACAAAGTTTAACTCAGA 6771
Db 943 IleCysArgLysLeuLysLeuAspProPheLeuThrProTyThrLysIleAsnSerArg 962
QY 6772 TGAATTAAGACTTAAATATAGACATAAAACCAATAAAACCCA-GAAGAAACCTAGGC 6830
Db 963 TrpIleLysAspLeuAsnValLysProLysThrIleLysThrLeuGluGluAsnLeuGly 982
QY 6831 AATACCATTGAGTATGACATGGGCAAGACTTCATGACTTAAACCAACCAAAAGCAATG 6890
Db 983 IleThrIleGlnAspIleGlyValGlyLysAspPheMetSerLysThrProLysAlaMet 1002
QY 6891 GCAACAAAGCCAAATAGACAAAGTGGGATCTGATTAACATATAGAGCTTCGACAGCA 6950
QY 1003 AlaThrLysAspLysIleAspLysTrpAspLeuIleLysLeuLysSerPheCysThrAla 1022
QY 6951 AAAAAAACTGTCATCAGAGTGAACAAGCAACCTACAGAATGGAGAAAATTTTGC AAT 7010
Db 1023 -LysGluThrThrIleA-gValAsnArgGlnProThrThrTrpGluLysIlePheAlaTh 1042
QY 7011 CTATCGATCTCACAAGGCTAATATATCCAGAGATCTACGAAGAACTTAAACAATTTACAA 7070
Db 1042 rTySerSerAspLysGlyLeuIleSerArgIleTyAsnGlnLeuLysGlnIleTyL 1062
QY 7071 GAAAAA----AACACCCCGTCAAAATATGGCAAGGATATGAGCAGACACTTCTCAAA 7126
Db 1062 sLysLysThrAsnAsnProIleLysLysTrpAlaLysAspMetAsnArgHisPheSerL 1082
QY 7127 AGAAGACATTTATGAGCCCAACAAACATATGAAAAAACCTCATCATCATTCGTCGTAG 7186
Db 1082 sGluAspIleTyAlaAlaLysLysHisMetLysLysCysSerSerLeuAlaIleAr 1102
QY 7187 AGAAATGCAAAACAAAACACAGTGCATACATCATCTCATGTAGTAGATGGTGATCAC 7246
Db 1102 gLumMetGlnIleLysThrThrMetArgTyHisLeuThrProValArgMetAlaIle1 1122
QY 7247 TAAAAAGTCAGAAACAAACAATGCTGGAGAGATGTGGAGAAATAGGACACTTTTCCA 7306
Db 1122 eLysLysSerGlyAsnAsnArgCysTrpArgGlyCysGlyGluIleGlyThrLeuLeuHi 1142
QY 7307 CTGTTGGTGGGAATGTAATATTAGTCAACATTTGGGAAGACAGTGTGGAGATTCTCTAA 7366
Db 1142 scTyTrpAspCysLysLysLeuValGlnProLeuTrpLysSerValTrpArgPheLeuAr 1162
QY 7367 GGATAGAACACAGAAATATCATTTGACCCCAATCCCATCTACTAGTATATATACCCAAA 7426
Db 1162 gAspLeuGluLeuGluIleProPheAspProAlaIleProLeuLeuGlyIleTyProAs 1182
QY 7427 GGAATATAAATCATCTTATTATAAAGACACATGCACACATATGTTTATTGAGCAGCTGAT 7486
Db 1182 nGluTyLysSerCysCysTyLysAspThrCysThrArgMetPheIleAlaLeuPh 1202
QY 7487 CACAATAGCAAGACTTGGAAACCAACCAATGTCCATGCAGTGATGAGACTGATAAGAA 7546
Db 1202 eThrIleValLysThrTrpAsnGlnProLysCysProThrMetIleAspTrpIleLysL 1222
QY 7547 AACATGGCACATATACACCATGAAATACTATGACGCCATATAAAAG-GATGAGTTCATGTC 7605
Db 1222 sMetTrpHisIleTyThrMetGluTyTyAlaAlaIleLysAsnAspGluPheIleSe 1242
QY 7606 CTTTCAGAGATATGAGTGAAGCTGGAACCATCATCTCTCAGCAAACTAACACAGAACA 7665
Db 1242 rPheValGlyThrTrpMetLysLeuGluThrIleIleLeuSerLysLeuSerGlnGluG 1262
QY 7666 GAAACCAACCAACACATCTCTCACTTGTAAAGTGGAGT 7705
Db 1262 nLysThrLysHisArgIlePheSerLeuIleGlyGlyAsn 1275
RESULT 14
000362 PRELIMINARY; PRT; 1275 AA.
AC 000362;
DT 01-JUL-1997 (TRENBLrel. 04, Created)
DT 01-JUL-1997 (TRENBLrel. 04, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Putative pl50.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97285120; PubMed=9140393;
RA Sassaman D.M., Dombroski B.A., Moran J.V., Kimberland M.L., Naas T.P.,
RA DeBerardinis R.J., Gabriel A., Swergold G.D., Kazazian H.H. Jr.;
RT "Many human LI elements are capable of retrotransposition.";
```

RL Nat. Genet. 16:37-43(1997).
DR EMBL: U93564; AAC51263.1; .
DR InterPro: IPR005135; Exo_endo_phos.
DR InterPro: IPR000477; RVTse.
DR Pfam: PF03372; Exo_endo_phos; 1.
DR Pfam: PF00078; rvt; 1.
KW RNA-directed DNA polymerase.
SQ SEQUENCE 1275 AA; 149201 MW; 23D516D6B4358F28 CRC64;

Alignment Scores:
Pred. No.: 0 Length: 1275
Score: 5150.00 Matches: 1053
Percent Similarity: 78.69% Conservative: 77
Best Local Similarity: 73.33% Mismatches: 140
Query Match: 9.84% Indels: 173
Dbs: 4 Gaps: 7

US-10-083-853B-2 (1-29921) x 000362 (1-1275)

Qy 3428 ATGGCAGGATCAATTCACACATAATAATTAACCTTAATGTAATGGCTAAATCC 3487
Dbb 1 MetThrGlySerAsnSerHisIleThrIleLeuThrLeuAsnIleAsnGlyLeuAsnSer 20
Qy 3488 CCAATTAAAGACACAGACTGGCAATTTGGATAAAGAGTCAAGACCCATCAGTGTCTGT 3547
Dbb 21 AlaIleLysArgHisArgLeuAlaAsnTrpIleLysSerGlnAspProSerValCysCys 40
Qy 3548 ATTGAGGAGCCCATCTCACATGAAAGACACACATAGGCTCAAAATAAAGGATGAGG 3607
Dbb 41 IleGlnGluThrHisLeuThrCysArgAspThrHisArgLeuLysIleLysGlyTrpArg 60
Qy 3608 AAGATTACCAAGTAAATGAAACAAACAAAAAGAGGGGTGCAATCCTAGTCTCT 3667
Dbb 61 LysIleTrpGlnAlaAsnGlyLysGln--LysLysAlaGlyValAlaIleLeuValSer 79
Qy 3668 GATAAAGACAGCTTTAAACCAACAAAGATCAAAAGACAAAGAGGCCATTACATAATG 3727
Dbb 80 AspLysThrAspPheLysProThrLysIleLysArgAspLysGluGlyHisTyrIleMet 99
Qy 3728 GTAAAGCATCAATGGACAGAGAGCTAAGTATCTTAATATACATGACCCCAATACA 3787
Dbb 100 ValLysGlySerIleGlnGlnGluLeuThrIleLeuAsnIleTyrAlaProAsnThr 119
Qy 3788 GGAGCCACAGATCTAAGACAGAGTCTTAGAGACTCAAAAGAGACTTTGACTCCAC 3847
Dbb 120 GlyAlaProArgPheIleLysGlnValLeuSerAspLeuGlnArgAspLeuAspSerHis 139
Qy 3848 ACAATAATAGTGGAGTCTAAATAATAATAGACACTTTAACACCCCACTGCCAATATTA 3907
Dbb 140 ThrLeuIleMetGly-----AspPheAsnThrProLeuSerThrLeu 153
Qy 3908 GGAGATCAATGACAGAGAAATTAACAGGATATCCAGAGTGAAGTGAAGTCTGGAC 3967
Dbb 154 AspArgSerThrArgGlnLysValAsnLysAspThrGlnGluLeuAsnSerAlaLeuHis 173
Qy 3968 CAAGCGGACCTAATAGATATCTACAGACATCCCAACCCCAACCAACAGAAATATACACTC 4027
Dbb 174 GluAlaAspLeuIleAspIleTyrArgThrLeuHisProLysSerThrGluTyrThrPhe 193
Qy 4028 TTCTGACATCAGATACACTATTTTAAATGACATGATTAATTAAGTAAACACTC 4087
Dbb 194 PheSerAlaProHisHisThrTyrSerLysIleAspHisIleValGlySerLysAlaLeu 213
Qy 4088 CTCAGCAAAATGCAAAAGACAGAAATCTTAACAAACAGTCTCTCAGACTACAGTGCATC 4147
Dbb 214 LeuSerLysCysLysArgThrGluIleIleThrAsnTyrLeuSerAspHisSerAlaIle 233
Qy 4148 TATTAGAACTCAGAAATTAAGAACTCAGTCAAAATCACAACTACATGAAACTGAAC 4207
Dbb 234 LysLeuGluLeuArgIleLysAsnLeuThrGlnSerArgSerThrThrTrpLysLeuAsn 253
Qy 4208 AACCTGCTCTGAATGACTACTGGGTAAATAACAAATGAAGGCAAAATPAAGATGTC 4267
Dbb 4267

Dbb 254 AsnLeuLeuLeuAsnAspCysTrpValHisAsnGluMetLysAlaGluIleLysMetPhe 273
Qy 4268 TTTGAACCAATGAGACAAAGACACAATCTACAGAAATCTCTGGGCGCATATTTAAAGCA 4327
Dbb 274 PheGluThrAsnGluAsnLysAspThrThrTyrGlnAsnLeuTrpAspAlaPheLysAla 293
Qy 4328 GTGTGTAGAGGAAATTTATAGCAGCTAGATGCGCTACAAAGAGAAAGACAGGAAATTTAAA 4387
Dbb 294 ValCysArgGlyLysLeuIleAlaLeuAsnAlaTyrLysArgLysGlnGluArgSerLys 313
Qy 4388 ATAGACACCTTAACATCACAAATTAAGAACTAGAGAAAGAGAGCAAAACAATTTCAAA 4447
Dbb 314 IleAspThrLeuThrSerGlnLeuLysGluLeuLysGlnGluThrHisSerLys 333
Qy 4448 GCTAGCAGAGACAAAGAAATACTAAGATCAGACAGCAACTGAGGAGATAGAGACAAA 4507
Dbb 334 AlaSerArgArgGlnGluIleThrLysIleArgAlaGluLeuLysGluIleGluThrGln 353
Qy 4508 AAAAGCCCTTCAAAATAAATCAATCAATCCAGAGCTGTTTTTTGAAAAGATACGCAAAAT 4567
Dbb 354 LysThrLeuGlnLys-IleAsnGluSerArgSerTrpPheGluArgIleAsnLysIle 373
Qy 4568 ----AGACCACTAGACAGACTAATAAGAAAGAAAGAGAGAAATCAAGAGATGCAAT 4623
Dbb 373 eAspArgProLeuAlaArgLeuIleLysLysLysArgGluLysAsnGlnIleAspThrIle 393
Qy 4624 AAAAATGATAAAGGGGATATCACACCGATCCACAGAAATCAACAATATTATTCAGAGA 4683
Dbb 393 eLysAsnAspLysGlyAspIleThrAspProSerGluIleGlnThrThrIleArgGln 413
Qy 4684 ATATTATAACACCTCTATGCAAAATAAATAAGAAATCTAGAGAAATGGATAAATTCCT 4743
Dbb 413 uTyrTyrLysHisLeuTyrAlaAsnLysLeuGluAsnLeuGluMetAspThrPheLe 433
Qy 4744 GGACACATATGTAGCCCTGTATGACCTTGGGGGACAGACAAAGGGGGTGAATGCAGAA 4803
Dbb 433 uAsp-----434
Qy 4804 ATAAAGACAAAGACAAAGAGATATGTTTGAAGTAGGGCTAGGGGGCAACTTGCCTCT 4863
Dbb 434
Qy 4864 AATGGACAGGGCCCTGAGCTTTACACACCCCTCTGTATTATTATAGGCAAAAGAGATGC 4923
Dbb 434
Qy 4924 GAGAGGTGAGTTGGAGAAGAGAGGTGAGCTTTAGGTCCAGAGTAGGCCCTGCAAGACTGC 4983
Dbb 434
Qy 4984 ATTCCTCAAAACAATAGCTCTAGATGTCCAGTAGATAACCTCAAGGAGGAGTGCAGG 5043
Dbb 434
Qy 5044 GAGTGTAGGCCCTCAGCAAACTTCTAGGGCAGGCACAGAAAGTAAGTTTGGCCACATTCT 5103
Dbb 434
Qy 5104 GTATTACAGATAAACAGTTTGTGTTGATCAAGTAGCCCTCAGTGAATGCTGAGTTGG 5163
Dbb 434
Qy 5164 TCATGATCCCTTTGGCCCTTTTGGCTCCCAAAACACATACACCTCTCAAGACTTAAACCA 5223
Dbb 435 -----ThrTyrThrLeuProArgLeuAsnGln 443
Qy 5224 GGAGAGTCAAAATCCCTGAATATATACCAAGTAAACAAGTTCTAAATTTGAAGCAGTAAATGA 5283
Dbb 443 nGluGluValGluSerLeuAsnArgProIleThrGlySerGluIleValAlaIleIleAs 463
Qy 5284 TAGCCTACCAACCAAAAGAGTCCAGGACAGGAGTTCACAGCCAAATTTACACAGAG 5343
Dbb 463 nSerLeuProThrLysLysSerLeuGlyProAspArgPheThrAlaGluPheTyrGlnAr 483

QY	5344	GTACAAAGAGAGAGCTGGTACTATTCTCTGTGAAACTATTCCAAAAATAGAA-----AATGG	5400
Db	483	gTyLySgLUgLUValProPheLeuLeuLysLeuPheGlnSerIlleGluLySgLU-G	503
QY	5401	GAATCTCTCCTAACTCAATTTTACGAGGCCAGGCATCTCTGATACCAAAACCTAGCAGTG	5460
Db	503	lytleLeuProasnSerPheTyrgLUaLSerIlleLeuIlleProLySproGlyargA	523
QY	5461	ACACACAAAAAGAGAAATTCAGGCCCATCTCCTGATCAACATTGATGTGAAAATCC	5520
Db	523	spThrThryLySgLUasnPheArgProIlleSerLeuMetasnIlleaspAlaLySleL	543
QY	5521	TCAATAAAATACTGGCAACCACTCCAGCAGCACATCAAAAAGCTTATCTACCATGATC	5580
Db	543	euAsnLySleLeuAlaAsnArgIlleGlnProIlleLySleLeuIlleHisAspG	563
QY	5581	AGTTGGCTCATCTCTGGATGCAAGCTGGTTCAAAATATGCAATCAATAATGTAG	5640
Db	563	lnValGlyPheIleProGlyMetGlnGlyTrpPheAsnIlleArgLySleSerIlleAsnValI	583
QY	5641	GCCATCACATAACAGAACCAATGACAAAAACACATGATTATCTCAATAGATCGAGAA	5700
Db	583	leclnHisIleasnArgAlaLySaspLySasnHisMetIlleIleSerIleaspAlaGluL	603
QY	5701	AGGCCTTTGTCAAAATTCACAGCCCTTCATCGCTAAAAAATTCCTCAGTAAACTAGGTATCG	5760
Db	603	ysAlaPheaspLySleGlnProPheMetLeuLySthryLeuAsnLySleuGlyIleA	623
QY	5761	ATGGAAATGATCTCAAAATATAAGAGCTATTATTATAC--AAACCCACAGCCAAATATCATAC	5819
Db	623	spGlyThryrPheLySleIleleArGAlaIleTyraspLySproThrAlaAsnIleIleL	643
QY	5820	TGAATGGCAAAACTGAGCATCTCCCTTTGAGAACTGGCCACAGACAGGATGCCCTC	5879
Db	643	euAsnGlylnLySleuGluLaPheProLeuLySthryGlyThryArgGlnGlycysProL	663
QY	5880	TCTCACCACTCTATTCAAGTACTATTGGAAGTCTGGCCAGGCCAATCAGGCAATAGA	5939
Db	663	euSerProLeuLeuPheasnIleValLeuGluValLeuAlaArgAlaIleargGlnGluL	683
QY	5940	AGAATAAAGGGTATTCAATAGAAAGAGAGAGGATCATATTCTCTCTGTTTGCAGATG	5999
Db	683	ysGluIleLySgLyIleGlnLeuGlyLySgLUValLySleuSerLeuPheAlaAspA	703
QY	6000	ACATGTTGTATATTAGAAACCCCATCGTCTCAGGCCCAAAACTCTTTAACTGATAA	6059
Db	703	spMetIleValTyrrLeuGluasnProIleValSerAlaGlnAsnLeuLySleuIleS	723
QY	6060	GCAACTTCAGCAAGCTCTCAGCACAAAAATCAATGTGCAAAAATTCACAGCACTTCTTAT	6119
Db	723	erAsnPheSerLySValSerGlyTyrrLySleasnValGlnLySserGlnAlaPheLeuT	743
QY	6120	AGCCCAATAATAGACAAACAGAGAGCCAAATCATGATGTAACATTCATTCACAAATTGCTA	6179
Db	743	yrThrAsnAsnArglnThrGluSerGlnIleMetGlyIleGluLeuProPheThryIleAlaS	763
QY	6180	CAACAGATAAATACCTAGGAATACAACTTACAAGGCACCGCTAGGAATCTTCAAGG	6239
Db	763	erLyArgIleLySthryLeuGlyIleGlnLeuThryArgaspValLySaspLeuPheLySg	783
QY	6240	AGAACTACAAACCACTGATCAGGAAATAAGAGAGGACACAAACAAATGAAAAACATTC	6299
Db	783	luAsnTyrrLySproLeuLeuLySgLUllyLySgLUaspThryasnLySthryLySasnIleP	803
QY	6300	CATGCTCACAGATA-----GTAAAGATCATGAAAAATGCC-ATACTGCCCAAAAGTAAT	6352
Db	803	roCysSerTrpValGlyArgIleasnIleMetLySmetAlaIleLeuProLySValIleI	823
QY	6353	ATAGATTCAGTGCACCCCATCAAGCTACCATGACTTCTTTCACACAGAATTCGAAAAA	6412
Db	823	yrArgPheAsnAlaIleProIleLySleuProMetThryPheThryGluLeuGluLyS	843


```
Db 1202 eThrIleAlaLysThrTrpAsnGlnProLysCysProThrMetIleAspTrpIleLysLy 1222
QY 7547 AACATGCCATATACACCACTAAATCTACTGAGCCATATAAAG-CATGAGTTCATGTC 7605
Db 1222 sMetTrpHisIleYrThrMetGluTrpYrAlaAlaIleLysAsnAspGluPheIleSe 1242
QY 7606 CTTTGCAGAGATATGGTGAAGCTGGAAACCATCTCTCAGCAAACTAACACAAAGAACA 7665
Db 1242 rPheValGlyThrTrpMetLysLeuGluThrIleLeuSerLysLeuSerGlnGluGl 1262
QY 7666 GAAACCAACACACATGTTCTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 7705
Db 1262 nLysThrLysHisArgIlePheSerLeuIleGlyGlyAsn 1275

RESULT 15
O00372
ID O00372 PRELIMINARY; PRT; 1275 AA.
AC O00372;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative p150.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97285120; PubMed=9140393;
RA Sussman D.M., Dombroski B.A., Moran J.V., Kimberland M.L., Naas T.P.,
RA DeBerardinis R.J., Gabriel A., Swergold G.D., Kazazian H.H. Jr.;
RT "Many human LI elements are capable of retrotransposition.";
RL Nat. Genet. 16:37-43(1997).
DR EMBL; U93570; AAC51273.1; -;
DR InterPro; IPR005135; Exo_endo_phos.
DR DR Pfam; PF03372; Exo_endo_phos; 1.
DR DR Pfam; PF00078; rvt; 1.
DR RNA-directed DNA polymerase.
SQ SEQUENCE 1275 AA; 149111 MW; 4711B3BC22F767AE CRC64;

Alignment Scores:
Pred. No.: 0 Length: 1275
Score: 5133.00 Matches: 1045
Percent Similarity: 78.55% Conservative: 83
Best Local Similarity: 72.77% Mismatches: 142
Query Match: 9.81% Indels: 173
DB: 4 Gaps: 7

US-10-083-853B-2 (1-29921) x O00372 (1-1275)
QY 3428 ATGGCAGGATCAATTCACACATAATATTAACCTTAATGTAATGGCTAAATTC 3487
Db 1 MethrGlySerAsnSerHisIleThrIleLeuThrLeuAsnIleAsnGlyLeuAsnSer 20
QY 3488 CCAATTAAGACACACACTGGCAAAATTTGGATAAAGAGTCAAGACCCCATCAGTGTGCT 3547
Db 21 AlaIleLysArgHisArgLeuAlaSerTrpIleLysSerGlnAspProSerValCysCys 40
QY 3548 ATTGAGGAGCCCATCTCATGAAAGACACACATAGGCTCAAAATATAAGGGATGGAGG 3607
Db 41 IleGlnGluThrHisLeuMetCysArgAspThrHisArgLeuLysIleLysGlyTrpArg 60
QY 3608 AAGATTACCAAGTAATGAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 3667
Db 61 LysIleYrGlnAlaAsnGlyLysGln---LysLysAlaGlyValAlaIleLeuValSer 79
QY 3668 GATAAAACAGACTTTAAACCAACAAAGATCAATCAAAAGACAAAGAGCCCATCATATG 3727
Db 80 AspLysThrAspPheLysProThrLysIleLysArgAspLysGluGlyHisIleMet 99
QY 3728 GTAAAGGCATCAATGGAAACAAGAGAGCTAACTATCCTAAATATACATGACCCCAATACA 3787
```

```
Db 100 ValLysGlySerIleGlnGlnGluLeuThrIleLeuAsnIleYrAlaProAsnThr 119
QY 3788 GGAGCACCCAGATTCTATAAGACAGTCTTAGAGACCTACAAAGAGACTTTGACTCCAC 3847
Db 120 GlyAlaProArgPheIleLysGlnAspLeuSerAspLeuGlnArgAspLeuAspSerHis 139
QY 3848 ACAATAATAGTGGGAGTCTAAATAATAATAGACACTTTAACACCCCACTGCCAATATTA 3907
Db 140 ThrLeuIleMetGly-----AspPheAsnThrProLeuSerThrLeu 153
QY 3908 GGCAGATCAATGACAGAGAAAATTAACAGATATCCAGAGTTGAACTGAGCTCTGGAC 3967
Db 154 AspArgSerThrArgGlnLysValAsnLysAspThrGlnGluLeuAsnSerAlaLeuHis 173
QY 3968 CAAGCCGACCTTAATAGATATCTACAGAACTCCCAACCCCACTAACTAACAGATATATACACTC 4027
Db 174 GlnAlaAspLeuIleAspIleYrArgThrLeuHisProLysSerThrGluTrpPhe 193
QY 4028 TTCTCAGCATCACATTACACTATTTTAAATTTGACCATGTAATTTTAAAGTAAACACTC 4087
Db 194 PheSerAlaProHisHisThrYrSerLysIleAspHisIleLeuGlySerLysAlaLeu 213
QY 4088 CTCAGCAAAATGCAAAAGACAGAAATCCTTAACAAACAGTCTCTCAGACTACAGTGCATC 4147
Db 214 LeuSerLysCysLysArgThrGluIleThrAsnTyrLeuSerAspHisSerAlaIle 233
QY 4148 TATTAGAACTCAGAAATAGAACTCAGTCAAAATCACAATCACAATCAGTCAAACTGAACTGAAC 4207
Db 234 LysLeuGluLeuArgIleLysAsnLeuThrGlnSerArgSerThrTrpLysLeuAsn 253
QY 4208 AACCTGCTCTGATGACTACTGGTAAATAACAAAATGAAGCGCAAAATAAAGATGTC 4267
Db 254 AsnLeuLeuLeuAsnAspTyrTrpValHisAsnGluMetLysAlaGluIleLysMetPhe 273
QY 4268 TTGAAACCAATGAGACAAAGACACAATGTACCAGAACTCTCTGGGCGCATATTTAAAGCA 4327
Db 274 PheGluThrAsnLysAsnLysAspThrThrTyrGlnAsnLeuTyrAspThrPheLysAla 293
QY 4328 GTGTGTAGAGGGAATTTATAGCACTAGATGCTCTACAGAGAGAGAGAGAGAGAGAGAGAG 4387
Db 294 ValCysArgGlyLysPheThrAlaLeuAsnAlaTyrLysArgLysGlnGluArgSerLys 313
QY 4388 ATAGACACCTTAACATCACAAATTAAGAACTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4447
Db 314 IleAspThrLeuThrSerGlnLeuLysGluLeuLysGlnGlnGlnThrHisSerLys 333
QY 4448 GCTAGCAGACAGCAAGAAATTAACATCAGACGACGAACTGAAGAGAGATAGACACAA 4507
Db 334 AlaSerArgArgGlnGluIleThrLysIleArgAlaGluLeuLysGluIleGluThrGln 353
QY 4508 AAAGCCCTTCAAAATAATCAATGAATCCAGGAGCTGTTTTCGAAAGAGATCAGCAAAAT 4567
Db 354 LysThrLeuGlnLys-IleAsnGluSerArgSerTrpPheGluArgIleAsnLysIle 373
QY 4568 ----AGACCCTACAGACACTAATAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4623
Db 373 eAspArgProLeuAlaArgLeuIleLysLysLysArgGluLysAsnGlnIleAspThrIle 393
QY 4624 AAAAATGATAAGGGGATATCACCCGATCCACAGAAATACAAACTATTATTCAGAGA 4683
Db 393 eLysAsnAspLysGlyAspIleThrThrAspProThrGluIleGlnThrIleArgGly 413
QY 4684 ATATTATAACACCTCTATCAATTAACACTAGAAAATCTAGAGAAATGATAAATTCCT 4743
Db 413 uTyrTyrLysHisLeuTyrAlaAsnLysLeuGluAsnLeuGluGluMetAspLysPheLe 433
QY 4744 GGACACATATGTAGCCCTGTATGGACCTTGGGGGACAGACAAAGAGGGGTGAATGCAGAA 4803
Db 433 u----- 433
QY 4804 ATAAAGACAAAGACAAAGAGATGTTTGGAAAGTAGGGGTGAGGGGGCAACTTGCCTCT 4863
```

Db	433	-----	433
Qy	4864	AATGGACAAGGCCCTGAGCTTTACACCACCTCTGTATTTATTAGGCAAAAGAGATAGC	4923
Db	433	-----	433
Qy	4924	GAGAGGTGAGTTGGGAAGAAGAGGTACAGTGTTCAGAGTAGGTCCAGAGTAGGCCCTGCAAGACTGC	4983
Db	433	-----	433
Qy	4984	ATTCTCTCAACAATAGSCTCTAGATGTCGCCAGTAGATAACCTCAAGGAGCCAGTGCCAGG	5043
Db	433	-----	433
Qy	5044	GAGTGATGGCCCTCAGCAAAACCTTCTAGGCGAGCGACAGAAGTAAGTTGCCACACATCT	5103
Db	433	-----	433
Qy	5104	GTATTCACGATAACAGCTTTGGCTGTTGTGATCAAGTAGCTCCAGTGGGAATGCTGAGTTGG	5163
Db	433	-----	433
Qy	5164	TCATGATCCCTTTGGCCCTTTTGGCTCCCAAAACACATACACCTCTCAGAGACTAAACCA	5223
Db	434	-----	434
Qy	5224	GGAAGAAGTCAAAATCCCTGAATATACAGTAACAGTCTTAAATTTGAAGCAGTAATTGA	5283
Db	443	nGluGluValGluSerLeuAsnArgProIleThrGlySerGluIleValAlaIleIleAs	463
Qy	5284	TAGCCTTACCACCAACCAAAAGTCCAGGACGAGGATTCACAGCCAAATCTTACGAGAG	5343
Db	463	nSerLeuProThrLysLysSerProGlyProAspGlyPheThrAspGluPheTyrglnAr	483
Qy	5344	GTACAAAGAGAAGCTGTACTATTCTCTGAACTATTCCAAAAATAGAA--AATGG	5400
Db	483	gTyrlsGluGluLeuValProPheLeuLeuLysLeuPheGlnSerIleGluLysGlu-G	503
Qy	5401	GAATCTCCCTAACTCATTTTACGAGGCCAGCATCTCTGTATACCAAACTAGCAGTG	5460
Db	503	lytleLeuProAsnSerPheTyrgluAlaSerIleLeuIleProLysProGlyArgA	523
Qy	5461	ACACAACAAAGAGGAATTTTCAGGCCATATCCCTGATGACATTTGATGTAATCC	5520
Db	523	spThrThrLysLysGluAsnPheArgProIleSerLeuMetAsnIleAspAlaLysIleL	543
Qy	5521	TCAATAAATACCTGGCAACCAATCCAGCAGCACATCAAAAAGCTTATCTACCATGATC	5580
Db	543	euAsnLysIleLeuAlaAsnGlnIleGlnHisIleLysLysLeuIleHisAspG	563
Qy	5581	AGTTGGGTATCCCTGGGATGCAAGGTGGTTCAAAATATATGCAAAATCAATAAATGTAG	5640
Db	563	lnValGlyPheIleProGlyMetGlnGlyTrpPheAsnMetCysLysSerIleAsnValI	583
Qy	5641	GCCATCACATAAACACAACTATGCAAAAACCCACATGATTTCTCAATAGATGCAAAA	5700
Db	583	leGlnHisIleAsnArgAlaLysAspLysAsnHisMetIleIleSerIleAspAlaGluL	603
Qy	5701	AGGCCCTTTGTCAAATTCACAGCCCTTCATGCTAAATAATTCACGTAACAGTAGGTATCG	5760
Db	603	ysAlaLeuAspLysIleGlnGlnProPheValLeuLysThrLeuAsnLysLeuGlyIleA	623
Qy	5761	ATGGAATGTATCTCAAAATAATAAGACTATTATATAC-AAACCCAGGCAATATATATAC	5819
Db	623	spGlyThrTyrlsPheLysIleMetArgAlaIleTyrlsAspLysProThrAlaAsnIleIleL	643
Qy	5820	TGAATGGCAAAACCTGGAGGATTCCTTTGAGAACTGGCAGACAGGATGCCCTC	5879
Db	643	euAsnGlyGlnLysLeuGluAlaPheProLeuLysThrGlyThrArgGlnGlyCysProL	663
Qy	5880	TCTCACCCTCTATTCAAGACTATTGGAAGTTCTGGCCAGGGCAATTCAGGCATAGAGA	5939
Db	663	euSerProLeuLeuPheAsnIleValLeuGluValLeuAlaArgAlaIleArgGlnGluL	683
Qy	5940	AAGAAATAAAGGGTATTCAAAATAGAAAGAGGAGAGTATATGCTCTCTGTTTCGAGATG	5999
Db	683	ysGluIleLysGlyIleGlnLeuGlyLysGluGluValLysLeuSerLeuPheAlaAspA	703
Qy	6000	ACATGTTTGTATATTAGAAAACCCCATCTCTCAGGCCAAAACCTCCTTAAGCTGATAA	6059
Db	703	spMetIleValTyrlsLeuGluAsnProIleValSerAlaGlnAsnLeuLysLeuIleL	723
Qy	6060	GCAACTTCACAAAGTCTCAGGACACAAATCAATGTGCAAAATCAACAGCACTTCATAT	6119
Db	723	erAsnPheSerLysValSerGlyTyrlsIleAsnValGlnLysSerGlnAlaPheLeuT	743
Qy	6120	ACGCCAATAATAGACAAACAGAGCCAAATCATGAGTGAACCTCTCATTCACAAATTCCTA	6179
Db	743	yrThrAsnAsnArgGlnThrGluSerGlnIleMetGlyGluLeuProPheThrIleAlaL	763
Qy	6180	CAAGAGATAAATACTAGGAATACAACTTCAAGGGACACAGTAGGAATCTTCAAGG	6239
Db	763	erLysArgIleLysTyrlsLeuGlyIleGlnLeuThrArgAspValLysAspLeuPheLysG	783
Qy	6240	AGAACTACAAACCACTGATCAAGGAAATAAGAGGAGGACACAAACAATAAGGAAACATTC	6299
Db	783	luAsnTyrlsProLeuLeuLysGluIleLysGluAspThrAsnLysIleAsnIleP	803
Qy	6300	CATGCTCACAGATAGTAAGAAATCAT-----GAAATGCCATCTGCCCAAGTAAAT	6352
Db	803	rocysSerTrpValGlyArgIleAsnIleValLysMetAlaIleLeuProLysValIleT	823
Qy	6353	ATGATTCAGTGTCTACCCCATCAAGCTACCATTCAGTCTTCTTCACAGAAATGGAATAA	6412
Db	823	yrArgPheAsnAlaIleProIleAsnLeuProMetThrPhePheThrGluLeuGluLysT	843
Qy	6413	CAACTTTAAATTTTCATATGGAACCAAAAAGAGCCACAGAGCCAGCAACCTTAAGC	6472
Db	843	hrThrLeuLysPheIleTrpAsnGln-LysArgAlaArgIleAlaLysSerIleLeuSer	862
Qy	6473	AAAAAGACAAAGCTGGAGTATCATCTACCTGACTTAAACTATATACTATAAGGCTACA	6532
Db	863	GlnLysAsnLysAlaGlyGlyIleThrLeuProAspPheLysLeuTyrlsAlaThr	882
Qy	6533	GTAAACAAACATGCTGGTACTGGTACCAAAACAGATATATACCAATAGGAACACAACA	6592
Db	883	ValThrLysThrValTrpTyrlsTrpTyrlsGlnAsnArgAspIleAsnArgThr	902
Qy	6593	GAGACCTCAGAATATAC-ACCTGCAATCTACATCCATCTGATCTTTTACAAAACCTGACAA	6651
Db	903	GluProSerGluIleMetProHisIleTyrlsAsnTyrlsMetIlePheAspLysProGluLys	922
Qy	6652	ACAACGAATGGAAGGATTCCTTATTTAATAATGGTGTGGAAAACTGGCTAGCC	6711
Db	923	AsnLysGlnTrpGlyLysAspSerLeuPheAsnLysTrpCysTrpGluAsnTrpLeuAla	942
Qy	6712	ATATGCAAGAACTGAACTGGATCCCTTCTTACACCTTATACAAAGCTTACTCAAGA	6771
Db	943	IleCysArgLysLeuLysLeuAspProPheLeuThrProTyrlsThrLysIleAsnSerArg	962
Qy	6772	TGAATTAAGACTTAAATATAAGACATAAAACCATAAAAACCCA-GAAGAAACCTTAGGC	6830
Db	963	TrpIleLysAspLeuAsnValArgProLysThrIleLysThrLeuGluGluAsnLeuGly	982
Qy	6831	ATATCATTTCAGATATGACATGGCAAGACTTTCATGACTTAAACACCAACAAAGCAATG	6890
Db	983	IleThrIleGlnAspIleGlyValGlyLysAspPheMetSerLysThrProLysAlaMet	1002
Qy	6891	GCAACAAAGCCAAATACACAAGTGGGATCTGATTAACTATAGAGCTTCTGCACAGCA	6950
Db	1003	AlaThrLysAlaLysIleAspLysTrpAspLeuIleLysLeuLysPheCysThrAlaI	1022
Qy	6951	AAAAAACTGTCTCAGAGTGAACCAAGCAACCTACAGAAATGGGAAAAATTTTTCAT	7010
Db	1023	-LysGluThrThrIleArgValAsnArgGlnProThrThrTrpGluLysIlePheAlaTh	1042

```
QY 7011 CTATCGATCTGACAAAGGCTTAATATCCAGAGATCTACGAGAAGAACTTAACAAATTTACAA 7070
Db      |||||
1042 rTyrSerSerAspGlyLeuIleSerArgIleTyrAsnGluLeuLysGlnIleTyrLy 1062
QY 7071 GAAAAA--AACAAACCCGCTAAAATATGGGCAAGGATATGAGACAGACTTCTCAA 7126
Db      |||||
1062 sLysLysThrAsnAsnProIleLysTrpAlaLysAspMetAsnArgHisPheSerLy 1082
QY 7127 AGAAGACATTATGCGACCAACAACATATGAAAAACCTCATCATTTGGTCGTTAG 7186
Db      |||||
1082 sGluAspIleTyrAlaAlaLysLysHisMetLysLysCysSerSerSerLeuAlaIleAr 1102
QY 7187 AGAAATGCAAAACAAACACAGTGCATACCATCTCATGCTAGTAGTAGATGCTGATCAC 7246
Db      |||||
1102 gGluMetGlnIleLysThrThrMetArgTyrHisHisThrProValArgMetAlaIleI 1122
QY 7247 TAAAAAGTACGAAACAACAAATGCTGGAGAGGATGTGGAGAATAGGACACTTTTCCA 7306
Db      |||||
1122 eLysLysSerGlyAsnAsnArgCysTrpArgCysGlyGluIleGlyThrLeuLeuH 1142
QY 7307 CTGTTGGTGGGAATGTAATTAAGTCAACCATTTGTGAACACAGTGTGGAGATTCTTAA 7366
Db      |||||
1142 sCysTrpTrpAspCysLysLeuValGlnProLeuTrpLysSerValTrpArgPheLeuAr 1162
QY 7367 GGATCTAGAACACAGAAATATCATTTGACCCAGCAATCCCATTTACTGAGTATATACCCAA 7426
Db      |||||
1162 gaspLeuGluLeuGluIleProPheAspProAlaIleProLeuLeuGlyIleTyrProLy 1182
QY 7427 GGAATATATCATTTCTATATATAAGACACATGCACACATATGTTTATGCCACACTGAT 7486
Db      |||||
1182 sasptYrLysSerCysCysTyrLysAspThrCysThrArgMetPheIleAlaLeuPh 1202
QY 7487 CACAATAGCAAGACTTGGAAACCAACCAATCTCCATCAGTCAGTCATAGCTGGATAAGAA 7546
Db      |||||
1202 eThrIleAlaLysThrTrpLysGlnProLysCysProThrMetIleAspTrpIleLysLy 1222
QY 7547 AACATGGCACATATACACCATGAATTAATCTATGCGCCATAAAAAAG-GATGAGTTTCATGC 7605
Db      |||||
1222 sMetTrpHisIleTyrThrMetGluTyrTyrAlaAlaIleLysAsnAspGluPheIleSe 1242
QY 7606 CTTTGCAGAGATATGATGAGCTGGAACCATCATCTCTCAGCAAACTACACAAAGAACA 7665
Db      |||||
1242 rPheValGlyThrTrpMetLysLeuGluThrIleIleLeuSerLysLeuSerGlnGluGl 1262
QY 7666 GAAACCAACACACACATGTTCTCACTTGTAACTGGGAGT 7705
Db      |||||
1262 nLysThrLysHisArgIlePheSerLeuIleGlyGlyAsn 1275
```

Search completed: July 4, 2003, 20:41:33
Job time : 1356 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 4, 2003, 19:01:07 ; Search time 186.5 Seconds
(without alignments)
13308.453 Million cell updates/sec

US-10-083-853B-2
Defect score: 53328
Sequence: 1 gtatgtgaagaagcctca.....caattctgtgaagaagtaa 29921

Scoring table: BLOSUM62 Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_SPOOL/US10083853/runat_03072003_093611_8078/app_query.fasta_1.30087
-DB=SwissProt_40 -OFMT=fastan -SUFFIX=rsr -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10083853 -CGEN=1.1.311 -runat_03072003_093611_8078 -NCPU=6 -ICPU=3
-NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5139	9.8	1259	1	LN1_HUMAN
2	3658.5	7.0	1260	1	LN1_NYCCO
3	3548.5	6.8	1300	1	POL2_MOUSE
4	2145	4.1	1259	1	LN1_HUMAN
5	1479.5	2.8	1260	1	LN1_NYCCO
6	1450.5	2.8	1308	1	POL2_MOUSE
7	443	0.8	1308	1	YTX2_XENLA
8	412	0.8	593	1	ALU7_HUMAN
9	393	0.7	593	1	ALU6_HUMAN
10	392	0.7	593	1	ALU7_HUMAN
11	387	0.7	379	1	YOL1_MOUSE
12	385	0.7	593	1	ALU6_HUMAN
13	381.5	0.7	591	1	ALU8_HUMAN
14	358	0.7	585	1	ALU5_HUMAN
15	348.5	0.7	585	1	ALU5_HUMAN
16	347	0.7	591	1	ALU1_HUMAN
17	345.5	0.7	591	1	ALU8_HUMAN
18	334	0.6	587	1	ALU2_HUMAN

c	19	328	0.6	587	1	ALU3_HUMAN	P39190 homo sapien
	20	320	0.6	587	1	ALU2_HUMAN	P39189 homo sapien
	21	316.5	0.6	591	1	ALU1_HUMAN	P39188 homo sapien
	22	308	0.6	587	1	ALU3_HUMAN	P39190 homo sapien
c	23	290	0.6	603	1	ALU4_HUMAN	P39191 homo sapien
	24	284	0.5	603	1	ALU4_HUMAN	P39191 homo sapien
	25	246.5	0.5	317	1	TLM_MOUSE	P17408 mus musculus
	26	241.5	0.5	1025	1	PO21_NASVI	Q03278 nasonia vit
	27	235	0.4	916	1	RTJK_DROME	P21328 drosophila
c	28	233	0.4	317	1	TLM_MOUSE	P17408 mus musculus
	29	223	0.4	1057	1	POLR_DROME	P16423 drosophila
c	30	221	0.4	418	1	YY1_HUMAN	P49646 homo sapien
	31	214.5	0.4	1222	1	YMH5_CAEEL	P34472 caenorhabdi
	32	211	0.4	711	1	PO22_PORJA	Q03274 popillia ja
	33	209.5	0.4	1790	1	USO1_YEAST	P25386 saccharomyc
	34	201	0.4	2230	1	GOG4_HUMAN	Q13439 homo sapien
	35	199.5	0.4	1957	1	YD86_SCHPO	Q10411 schizosach
	36	198.5	0.4	2663	1	CENE_HUMAN	Q02224 homo sapien
	37	194	0.4	944	1	NUF1_YEAST	P32380 saccharomyc
	38	194.5	0.4	1805	1	HMW2_MYCGE	P47460 mycoplasma
	39	193.5	0.4	1875	1	MLP1_YEAST	Q02455 saccharomyc
	40	193	0.4	2116	1	MY52_DICDI	P08799 dictyosteli
	41	186	0.4	2869	1	RBP1_PLAVB	Q00798 plasmodium
	42	182.5	0.3	1251	1	RBP2_PLAVB	Q00799 plasmodium
	43	181.5	0.3	1005	1	RA50_METJA	Q58718 methanococc
	44	179	0.3	1312	1	RA50_YEAST	P12753 saccharomyc
	45	178	0.3	1453	1	Y373_BOVIN	Q9TU23 bos taurus

ALIGNMENTS

RESULT 1
LN1_HUMAN
ID LN1_HUMAN STANDARD; PRT: 1259 AA.
AC P08547;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-AUG-1988 (Rel. 08, Last annotation update)
DE LINE-1 reverse transcriptase homolog.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RX MEDLINE=86230917; PubMed=2423883;
RA Hattori M., Kuhara S., Takenaka O., Sakaki Y.;
RT "L1 family of repetitive DNA sequences in primates may be derived
RT from a sequence encoding a reverse transcriptase-related protein.";
RL Nature 321:625-628(1986).
CC -!- MISCELLANEOUS: THIS SEQUENCE WAS CONSTRUCTED FROM AN ALIGNMENT OF
CC PUBLISHED AND UNPUBLISHED SEQUENCES, DETERMINED IN VARIOUS
CC LABORATORIES, BELONGING TO THE LINE-1 FAMILY.
DR PIR: A25313; GNHUL1.
DR InterPro: IPR005135; Exo_endo_phos.
DR InterPro: IPR000300; IPPC.
DR InterPro: IPR000477; RVTse.
DR Pfam: PF000078; rvt; 1.
DR Pfam: PF03372; Exo_endo_phos; 1.
DR SMART: SM00128; IPPC; 1.
KW RNA-directed DNA polymerase.
SQ SEQUENCE 1259 AA; 147112 MW; F3BC18A3803919F5 CRC64;

Alignment Scores:
Pred. No.: 0 Length: 1259
Score: 5139.00 Matches: 1050
Percent Similarity: 79.00% Conservative: 71
Best Local Similarity: 74.00% Mismatches: 132
Query Match: 9.82% Indels: 173
DB: 1 Gaps: 7

US-10-083-853B-2 (1-29921) x LN1_HUMAN (1-1259)

```
QY 3434 GGATCAATTCACACATAAATAATTAACCTTAATGCTAAATGGCTAAATCCCAATT 3493
Db 2 GlySerAsnSerHisIleThrIleLeuThrLeuAsnValAsnGlyLeuAsnAlaProIle 21
QY 3494 AAAAGACACAGACTGGCAAAATTTGGATAAAGAGTCAAGACCCATCAGTGTCTGATTTCAG 3553
Db 22 LysArgHisArgLeuAlaAsnTrpIleLysSerGlnAspProSerValCysCysIleGln 41
QY 3554 GAGGCCATCTCACATGAAAGACACACATAGGCTCAAAATAAAGGGATGGAGAGATT 3613
Db 42 GluThrHisLeuThrCysArgAspThrHisArgLeuLysIleLysGlyTrpArgAsnIle 61
QY 3614 TACCAAGTAATGAAACAAAAAAGAGGGGTGCCAATCTCTAGTCTCTGTATAAA 3673
Db 62 TyrGlnAlaAsnGlyLysGln---LysLysAlaGlyValAlaIleLeuValSerAspLys 80
QY 3674 ACAGACTTTAAACCAACAAGATCAAAAGACAAAGAGACAAAGAGCCATTACATAATGGTAAAG 3733
Db 81 ThrAspPheLysProThrLysIleLysArgAspLysGlyHisIleMetValLys 100
QY 3734 GCATCAATGGAACAAGAGCTAACTATCTTAATATATACATGCCCAATACAGGAGCA 3793
Db 101 GlySerIleGlnGlnGluLeuThrIleLeuAsnIleThrAlaProAsnThrGlyAla 120
QY 3794 CCGAGATTCAAAAGCAAGTTCTTAGAGACTCAAAAGAGACTTTGACTCCCAACAATA 3853
Db 121 ProArgPheIleLysGlnValLeuSerAspLeuGlnArgAspLeuAspSerHisThrIle 140
QY 3854 ATAGTGGAGCTCTAAATAAATAGACACTTTAAACACCCCACTGCCAATATTAGGCAGA 3913
Db 141 IleMetGly-----AspPheAsnThrProLeuSerThrLeuAspArg 154
QY 3914 TCAATGAGACAGAAAAATTAAACAGGATATCCAGAGTTGAACCTGAGCTCTGGACCAAGCG 3973
Db 155 SerThrArgGlnLysIleAsnLysAspIleGlnGluLeuAsnSerAlaLeuHisGlnAla 174
QY 3974 GACCTAATAGATATCTACAGAACTCCCAACCCCAATCAACAGAAATATACACTCTCTCA 4033
Db 175 AspLeuIleAspIleThrArgThrLeuHisProLysSerThrGluThrThrPhePheSer 194
QY 4034 GCATCATACACTATTTTAAATTTGACCATGTAATTTAAGTAAACACTCTCAGC 4093
Db 195 AlaProHisThrThrSerLysThrAspHisIleLeuGlySerLysThrLeuLeuSer 214
QY 4094 AAATGCAAAAGACAGAAATCTTAAACAAACAGCTCTCTCAGCTACAGTGAATCTATTTA 4153
Db 215 LysCysLysArgThrGluIleIleThrAsnCysLeuSerAspHisSerAlaIleLysLeu 234
QY 4154 GAACCTAGAAATTAAGAACTACTCAAAATCACAACTACATGGAACCTGAACACCTG 4213
Db 235 GluLeuArgIleLysLysLeuThrGlnAsnHisSerThrThrTrpLysLeuAsnAsnLeu 254
QY 4214 CTCCTGAATGACTACTGGGTAAATACAAATCAAGCAAAATAAAGATGTTCTTGA 4273
Db 255 LeuLeuAsnAspTrpTrpValHisAsnGluMetLysAlaGluIleLysLysPhePheGlu 274
QY 4274 ACCAATGACAACAAAGACAAATGTACCAGAAATCTCTGGGGCATATTTAAAGCAGTGTG 4333
Db 275 ThrAsnGluAsnLysAspThrThrThrGlnAsnLeuTrpAspThrAlaLysAlaValCys 294
QY 4334 AGAGGAAATTTATAGACTAGATGCTTCAAGAGAAAGCAGAAATATCTTAAATAGAC 4393
Db 295 ArgGlyLysPheIleAlaLeuAsnAlaHisLysArgLysGlnGluArgSerLysIleAsp 314
QY 4394 ACCTTAACATCACAATTTAAAGAACTAGAGAAAGACAAACAATTCARAAGCTAGC 4453
Db 315 ThrLeuIleSerGlnLeuLysGluLeuGluLysGlnGlnThrAsnSerLysAlaSer 334
QY 4454 AGAAGACAAGAAATAACTAAGATCACAGACAGAACTGAAGGAGATAGACACAAAAAGCC 4513
Db 335 ArgArgGlnGluIleIleLysIleArgAlaGluLeuLysGluIleGluThrGlnLysThr 354
QY 4514 CTTCAANTAAATCAATGAATCCAGGAGCTGGTGTGTTTTTGAAGAAATCAGCAAAAT 4569
Db 354 ystIleLeuAlaAsnGlnIleGlnGlnHisIleLysLysLeuIleHisAspGlnValG 564
Db 5587 GCGTCATCCTCGGATGCAAGCTGGTTCAAAATATGCAANTCAATAAATGTAGGCCATC 5646
QY 335 LeuGlnLys-IleAsnGluSerArgSerTrpPhePheGluLysIleAsnLysIleAspAr 374
QY 4570 ACCACTAGACAGACTAATAAAGAAAGAAAGAGAGAACTCAAGAGATCAATAAAAA 4629
Db 374 gProLeuAlaArgLeuIleLysLysArgGluLysAsnGlnIleAspThrIleLysAs 394
QY 4630 TGATAAAGGGGATATCACACCCGATCCACAGAAATACAAACTATTATCAGAGAATATTA 4689
Db 394 nAspArgGlyAspIleThrThrAspProThrGluIleGlnThrThrIleArgGluTrpTy 414
QY 4690 TAAACACCTCTATGCAAAATAAATACTAGAAAATCTAGAGAAATGGATAAATCTCTGGACAC 4749
Db 414 rLysHisLeuTyrAlaAsnLysLeuGluAsnLeuGluMetAspLysPheLeuAsp-- 433
QY 4750 ATATGTAGCTGTATGGACCTTGGGGACAGACAAAGAGGGGTGAATGCAGAAATAAAA 4809
Db 433 ----- 433
QY 4810 GACAAAGACAAAAAGAGTATGTTTGAAGTAGGGGTGAGGGGCAACTTGCCTCTAATGGA 4869
Db 433 ----- 433
QY 4870 CAAGGCCCTGAGCTTTACACACCCTCTGTATTATTATTAGGCAAAAGAGATAGCGAGG 4929
Db 433 ----- 433
QY 4930 GTGAGTTGGAAGAAGAGGTGAGTGTAGTCCAGAGTAGGCCCTGCAAGACTGCATTCCT 4989
Db 433 ----- 433
QY 4990 CAAACAATAAGGCTTAGATGTCCAGTAGATAACCTCAAGGAGCCAGTGCCAGGAGTGA 5049
Db 433 ----- 433
QY 5050 TGCCCTCAGCAAAACCTTCTAGGGCAGGCACAGAAAGTAAGTTTGGCCACATTCCTGTATTC 5109
Db 433 ----- 433
QY 5110 ACGATAACAGATTTGCTGTTGTATGATAGTCCAGTCCAGTGAATGCTGAGTTGGTCATGA 5169
Db 433 ----- 433
QY 5170 TCCTTTTGGCCTTTTGGCTCCCAAAACACATACACCCTCTCAAGACTTAAACCCAGGAAGA 5229
Db 434 -----ThrTyrThrLeuProArgLeuAsnGlnGluI 444
QY 5230 AGTCAAAATCCCTGAATATACAGTAACAAGTTCTTAAATTTGAAGCAGTAAATTCATAGCCT 5289
Db 444 uValGluSerLeuAsnArgProIleThrSerSerGluIleGluAlaIleIleAsnSerLe 464
QY 5290 ACCAACCAAAAAGTCCAGGACCCAGGATTCACGCCAAATTCACCAATTTCTACAGAGGTACAA 5349
Db 464 uProAsnLysLysSerProGlyProGlyPheThrAlaGluPheThrGlnArgTrpTy 484
QY 5350 AGAAGAGCTGGTACTATTCTCTGAACTATTCCAAAAAATAGAA---AATGGGAATCC 5406
Db 484 sGluGluLeuValProPheLeuLeuLysLeuPheGlnSerIleGluLysGlu-GlyIleL 504
QY 5407 TCCTTAATCTATTTTACGAGGCCAGCATCATCTGTATACAAACCTACAGTGCAGACAA 5466
Db 504 euProAsnSerPheThrGluAlaSerIleIleLeuIleProLysProGlyArgAspThrT 524
QY 5467 CAAAAAGAGAAATTTTCAGGCCCATATCCTGTATGAACATGATGTGAAAAATCCTCAATA 5526
Db 524 hrLysLysGluAsnPheArgProIleSerLeuMetAsnIleAspAlaLysIleLeuAsnL 544
QY 5527 AAATACTGCAACCAAAATCCAGCAGCATCAAAAAGCTTATCTACCATGATCAAGTTG 5586
Db 544 ystIleLeuAlaAsnGlnIleGlnGlnHisIleLysLysLeuIleHisAspGlnValG 564
QY 5587 GCGTCATCCTCGGATGCAAGCTGGTTCAAAATATGCAANTCAATAAATGTAGGCCATC 5646
Db 5587 GCGTCATCCTCGGATGCAAGCTGGTTCAAAATATGCAANTCAATAAATGTAGGCCATC 5646
```

Db	564	lyPheIleProAlaMetGInGlyTrpPheAsnIleArgLysSerIleAsnIleIleGlnH	584
QY	5647	ACATAAACAGAACCAATGACAAAAACACATGATTATCTCAATAGATGACAGAAAGGCCT	5706
Db	584	isIleAsnArgThrLysAspThrAsnHisMetIleIleSerIleAspAlaGluLysAlaP	604
QY	5707	TTGTCAAAATTCACAGCCCTTCATGCTCAAAAAATTCAGTAAACATAGGTATGATGGAA	5766
Db	604	heAspLysIleGInGInProPheMetLeuLysProLeuAsnLysLeuGlyIleAspGlyT	624
QY	5767	TGTATCTCAAAATAAATAGAGCTATTATTATAC-AAAGCCACAGCAATATCATCTAATG	5825
Db	624	hrTyrlLeuLysIleIleArgAlaIleTyrlAspLysProThrAlaAsnIleIleLeuAsnG	644
QY	5826	GGCAAAACTGGAGCAATCCCTTTGAGAACTGGCACAGCAAGGATGCCCTCTCTCAC	5885
Db	644	lyGlnLysLeuGluAlaProProLeuLysThrGlyThrArgGlnGlyCysProLeuSerP	664
QY	5886	CACCTCTATTCAAGATACATATTGGAAGTTCTTGCCAGGGCAATPACGGCAATAGAAAGAA	5945
Db	664	roLeuLeuProAsnIleValLeuGluValLeuAlaArgAlaIleArgGlnGluLysGluI	684
QY	5946	TAAAGGTATTCAATAGAAAGAGAGAGAGTCAATATGCTCTGTTGTCAGATGACATGT	6005
Db	684	leLysGlyIleGInLeuGlyLysGluGluValLysLeuSerLeuPheAlaAspMetI	704
QY	6006	TTGTATATTAGAAAACCCCATCTCTCAGGGCCAAAAACTCCCTTAAGCTGATAGCAACT	6065
Db	704	leValTyrlLeuGluAsnProIleValSerAlaGlnAsnLeuLysLeuIleSerAsnP	724
QY	6066	TCAGCAAAAGTCTCAGGACACAAAATCAATGTGCAAAAATCACAAAGCATTTCTTACGCCA	6125
Db	724	heSerLysValSerGlyTyrlLysIleAsnValGlnLysSerGlnAlaPheLeuTyrlThra	744
QY	6126	ATAATACACAAACAGAGACCAATCATGAGTGAACCTCATTCACAAATGCTACAAGA	6185
Db	744	snAsnArgGlnThrGluSerGlnIleMetSerGluLeuProPheThrIleAlaSerLysA	764
QY	6186	GAATAAAATACCTAGGAATACAACTTACAAAGGACACAGTAGGAACCTCTCAAGGAGAACT	6245
Db	764	rgIleLysTyrlLeuGlyIleGlnLeuThrArgAspValLysAspLeuPheLysGluAsnT	784
QY	6246	ACAAACCTGATCAAGGAATAAAGAGAGACACAAACAAATGGAAAAACATTCATGCT	6305
Db	784	ylLysProLeuLeuAsnGluIleLysGluAspThrAsnLysTrpLysAsnIleProCysS	804
QY	6306	CACAGATAGTAAGAATCAT-----GAAATGCCATACCTGCCCAAGTAATATAGAT	6358
Db	804	erTrpValGlyArgIleAsnIleValLysMetAlaIleLeuProLysValIleTyrlArgP	824
QY	6359	TCAGTGTACCCCATCAAGCTACCATGACTTCTTCAGAGAAATTTGGAAAAACAACCTT	6418
Db	824	heAsnAlaIleProIleLysLeuProMetThrPhePheThrGluLeuGluLysThrL	844
QY	6419	TAAATTTCAATATGGAACCAAAAAAGAGCCCAAGCCCAAGACAAATCTTAAGCAAAAG	6478
Db	844	eulysPheIleTrpAsnGln-LysArgAlaHisIleAlaLysSerThrLeuSerGlnLys	863
QY	6479	AACAAAGCTGGAGTATCATGCTACCTGACTTAAACTATATATACCTATACGCTACAGTAAC	6538
Db	864	AsnLysAlaGlyIleThrLeuProAspPheLysLeuTyrlTyrlLysAlaThrValThr	883
QY	6539	AAACCTGCATGGTACTGACCAAAAACAGATATATAGACCAATGGAAACAGACAGAGACC	6598
Db	884	LysThrAlaTrpTyrlTyrlGlnAsnArgAspIleAspGlnTrpAsnArgThrGluPro	903
QY	6599	TCAGAAATTAC-ACTGCAATCTACATCCATCTGATCTTTGACAAACCTGACAAAAACAAG	6657
Db	904	SerGluIleMetProHisIleTyrlAsnTyrlLeuIlePheAspLysProGluLysAsnLys	923
QY	6658	CAATGAAAAGGATCCCTATTATTAATGCTGTGGAAAAACCTGGCTAGCATATGC	6717
Db	924	GlnTrpGlyLysAspSerLeuPheAsnLysTrpCysTrpGluAsnTrpLeuAlaIleCys	943
QY	6718	AGAAAGCTGAAAGCTGATCCCTTCCCTTACACCTTATACAAAGCTTAACCTCAAGATCAAT	6777
Db	944	ArgLysLeuLysLeuAspProPheLeuThrProTyrlThrLysIleAsnSerArgTrpIle	963
QY	6778	AAAGACTTAATATAAGACATAAAACCAATAAAACCCA-GAAGAAACCTTAGCAATACC	6836
Db	964	LysAspLeuAsnValArgProLysThrIleLysThrLeuGluLysAsnLeuGlyAsnThr	983
QY	6837	ATTCAGATATGACATGGCAAGACTTCATCACTAAACACCAACCAACCAATGGCAACA	6896
Db	984	IleGlnAspIleGlyMetGlyLysAspPheMetThrLysThrProLysAlaMetAlaThr	1003
QY	6897	AAAGCAAAATAGACAAGTGGGATCTGATTAAACTATAGAGCTCTGCACAGCAAAAAA	6956
Db	1004	LysValLysIleAspArgTrpAspLeuIleLysLeuLysSerPheCysThrAla-LysG	1023
QY	6957	AACTGTCATCAGAGTGAACAGCAACCTACAGAAATGGGAAATTTTTGCAATCTATCG	7016
Db	1023	uThrThrIleArgValAsnArgGlnProThrLysTrpGluLysIlePheAlaIleTyrlSe	1043
QY	7017	ATCTGACAAAGGCTAATATCCAGAGATCTACGAAGAACTTAAACAAATTTACAGAAAAA	7076
Db	1043	rSerAspLysGlyLeuIleSerArgIleTyrlAsnGluLeuLysGlnIleTyrlLysLysL	1063
QY	7077	---AAACAACCCCTCAAAATATGGCAAGGATATGAGCAGACACTTCTCAAAAGAAGA	7132
Db	1063	sThrAsnAsnProIleLysLysTrpAlaLysAspMetAsnArgHisProSerLysGluAs	1083
QY	7133	CATTATGAGCCCAACAAATATGAAAAACCTCATCATCATTTGGTCTGTAGAGAAAT	7192
Db	1083	piLeTyrlAlaLysLysHisMetLysLysCysSerSerSerLeuAlaIleArgGluMe	1103
QY	7193	GCAAAACAAACCCACAGTGCATACCATCTCATGCTAGTAGTAGTGTGATGATCACTAAA	7252
Db	1103	tGlnIleLysThrThrMetArgTyrlHisLeuThrProValArgMetAlaIleIleLysL	1123
QY	7253	GTGAGAAACAAACAAATGCTGGAGAGATGTGGAGAAATAGGAACACTTTTCCACTGTTG	7312
Db	1123	sSerGlyAsnAsnArgCysTrpArgGlyCysGlyGluIleGlyThrLeuLeuHisCysTr	1143
QY	7313	GTGGGAATGTAATATTAGTCAACCATTTGGGAAGACAGTGTGGAGATTCCTTAAGGATCT	7372
Db	1143	pTrpAspCysLysLeuValGlnProLeuTrpLysSerValTrpArgPheLeuArgAspLe	1163
QY	7373	AGAACCAAGAAATCATTTGACCCCACTATCCATCTACGATATATACCCCAAGGAATA	7432
Db	1163	uGluLeuGluIleProPheAspProAlaIleProLeuLeuGlyIleTyrlProLysAspTy	1183
QY	7433	TAAATCATTTCTATTATAAGACACATGACACATATGTTTATTGACGACTGTATCAAT	7492
Db	1183	rLysSerCysCysTyrlLysAspThrCysThrArgMetPheIleAlaLeuPheThrI	1203
QY	7493	AGCAAGACTTGAACCAACCCCAATGTCCATCAGTGTAGACTGTGATTAAGAAACATG	7552
Db	1203	eAlaLysThrTrpAsnGlnProLysCysProThrMetIleAspTrpIleLysMetTr	1223
QY	7553	GCACATATACACCATGAAATCTATGACCCATATAAG-GATGAGTTCATGCTCCTTGC	7611
Db	1223	pHisIleTyrlThrMetGluTyrlTyrlAlaAlaIleLysAsnAspGluProMetSerProVa	1243
QY	7612	AGAGATATGATGAAGCTGGAACCATCATTTCTCAGCAAACTAACACA 7660	
Db	1243	IGlyThrTrpMetLysLeuGluThrIleIleLeuSerLysLeuSerGln 1259	
RESULT 2			
LINE1_NYCCO	LINE1_NYCCO	STANDARD;	PRT; 1260 AA.
ID	AC	P08548;	
DT	DT	01-AUG-1988 (Rel. 08, Created)	
DT	DT	01-AUG-1988 (Rel. 08, Last sequence update)	
DT	DT	01-AUG-1988 (Rel. 08, Last annotation update)	
DE	DE	LINE-1 reverse transcriptase homology.	


```
Db 931 ProGluMetAsnProHisThrTyrglyHisLeullePheAspLysGlyAlaLysThrIle 950
QY 6658 CAATGGAAGAGGATTCCTTATTAATAAATGGTGTGGAAGAACTGGCTAGCCATATGC 6717
Db 951 GlnTrpLysLysAspSerIlePheAsnAsnTrpCysTrpHisAsnTrpLeuSerCys 970
QY 6718 AGAAGCTGAAGAGGATTCCTTCTTACACCTTATACAAAAGTTAACTCAAGATGAA 6777
Db 971 ArgArgMetArgIleAspProTyLeuSerProCysThrLysValLysSerLysTrpIle 990
QY 6778 AAAGACITTAATATAAGACATAAACCAATAAACCCCA-GAAGAAACCTAGGCAATACC 6836
Db 991 LysGluLeuHisIleLysProGluThrLeuLysLeuIleGluLysValGlyLysSer 1010
QY 6837 ATTCAGGATATGACATGGGCAAGACTTCAATGACTAAACACCAAGCAATGGCAACA 6896
Db 1011 LeuGluAspMetGlyThrGlyGluLysPheLeuAsnArgThrAlaMetAlaCysAlaVal 1030
QY 6897 AAAGCAAAATAGACAAGTGGGATCTGATTAACTATAGAGCTTCTGCACAGCAAAAGAA 6956
Db 1031 ArgSerArgIleAspLysTrpAspLeuMetLysLeuGlnSerPheCysLysAla-LysAs 1050
QY 6957 AACTGTCTCATCAGAGTGAACAGCAACCTACAGATGGGAGAAATTTTTCGAATCTATCG 7016
Db 1050 pThrValAsnLysThrLysArgProProThrAspTrpGluArgIlePheThrTrpProLy 1070
QY 7017 ATCTGACAAAGGCTTAATATCCAGAGATCTACGAGAAGCTTAACAAATTTA----CAAGA 7072
Db 1070 sSerAspArgGlyLeulleSerAsnIleTyLysGluLeuLysValaspPheArgLy 1090
QY 7073 AAAAACAACCCCTCAAAATATATGGGAAAGATATGAGCAGACACTTCTCAAAAGAGA 7132
Db 1090 sSerAsnAsnProIleLysTrpGlySerGluLeuAsnLysGluPheSerProGluGl 1110
QY 7133 CATTTATCCAGCAACCAACATATCAAAAACCTCATCATCTGCTGTAGAGAAAT 7192
Db 1110 uTyArgMetAlaGluLysHisLeuLysLysCysSerThrSerLeulleIleArgGluMe 1130
QY 7193 GCAAAACAAACACACAGTGCATACCATCTCATCTAGTGTAGATGGTGATCACTAAAGAA 7252
Db 1130 tGlnIleLysThrThrLeuArgPheHisLeuThrProValArgMetAlaLysIleLysAs 1150
QY 7253 GTCAGGAACACAAATATCTGGAGAGGATGTGGAGAAATAGGAACACTTTTCCACTGTTG 7312
Db 1150 nSerGlyAspSerArgCysTrpArgGlyCysGlyGluArgGlyThrLeuLeuHisCysTr 1170
QY 7313 GTGGCAATGTAAATAGTTCACCAATGTGGAGACAGTGTGGAGATTCCTTAAGGATCT 7372
Db 1170 pTrpGluCysArgLeuValGlnProLeuTrpLysSerValTrpArgPheLeuArgLysLe 1190
QY 7373 AGAACCAAGAAATATCATTTGACCCAGCAATCCCATTTACTGAGTATATACCAAGGAATA 7432
Db 1190 uAspIleValLeuProGluAspProAlaIleProLeuLeuGlyIleTyProGluAspAl 1210
QY 7433 TAAATCATCTTATTAAGACACATGCACACATATGTTTATTTGCACACATGATCACAAT 7492
Db 1210 aProThr---GlyLysLysAspThrCysSerThrMetPheIleAlaLeuPheIleI 1229
QY 7493 AGCAAGACTTGGAAACCAACCAATGCTCCATCATGATGATAGCTGGATAAGAAACATG 7552
Db 1229 eAlaArgSerTrpLysGluProArgCysProSerThrGluGluTrpIleGlnLysMetTr 1249
QY 7553 GCACATATACACCATGAATATCTATGCAGCCAT-AAAAAGGATGAGTTTCATGCTCTTGC 7611
Db 1249 pTyIleTyThrMetGluTyTrSerAlaIleLysLysAsnGluPheMetLysPheLe 1269
QY 7612 AGACATATGATGAAGCTGGAACCACTATCTCAGCAAACTAACCAAGAACAGAGAAC 7671
Db 1269 uAlaLysTrpMetAspLeuGluGlyIleLeuSerGluValThrHisSerGlnArgAs 1289
QY 7672 CAACACCACTGTCTCACTTGTAACTGGG 7702
Db 1289 nSerHisAsnMetTyTrSerLeulleSerGly 1299
```

```
RESULT 4
LIN1_HUMAN
ID LIN1_HUMAN STANDARD; PRT; 1259 AA.
AC P08547;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-AUG-1988 (Rel. 08, Last annotation update)
DE LINE-1 reverse transcriptase homolog.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
MEDLINE=86230917; PubMed=2423883;
RA Hattori M., Kuhara S., Takenaka O., Sakaki Y.;
RT "Li family of repetitive DNA sequences in primates may be derived
RT from a sequence encoding a reverse transcriptase-related protein.";
RL Nature 321:625-628(1986).
CC -!- MISCELLANEOUS: THIS SEQUENCE WAS CONSTRUCTED FROM AN ALIGNMENT OF
CC PUBLISHED AND UNPUBLISHED SEQUENCES, DETERMINED IN VARIOUS
CC LABORATORIES, BELONGING TO THE LINE-1 FAMILY.
DR PIR; A25313; GNHULL.
DR InterPro; IPR005135; Exo_endo_phos.
DR InterPro; IPR000300; IPPC.
DR InterPro; IPR000477; RVTse.
DR Pfam; PF000078; Rvt; 1.
DR Pfam; PF03372; Exo_endo_phos; 1.
DR SMART; SM00128; IPPC; 1.
DR KW RNA-directed DNA polymerase.
SQ SEQUENCE 1259 AA; 147112 MW; F3BC18A3803919F5 CRC64;

Alignment Scores:
Pred. No.: 1.35e-146 Length: 1259
Score: 2145.00 Matches: 394
Percent Similarity: 95.07% Conservative: 11
Best Local Similarity: 92.49% Mismatches: 21
Query Match: 4.09% Indels: 0
DB: 1 Gaps: 0

US-10-083-853B-2 (1-29921) x LIN1_HUMAN (1-1259)
QY 29919 ACTTCTTCACAGAAATGGAAAACTACTTTAAAGTTCATGTGGAACCAAAAAGAGCC 29860
Db 834 ThrPhePheThrGluLeuGluLysThrThrLeuLysPheIleTrpAsnGlnLysArgAla 853
QY 29859 CACATCCCAAGTCAATCCTTAAGCCAAAGAAACAAAGCTGGAGGCATCTGCTACCTGAC 29800
Db 854 HisIleAlaLysSerThrLeuSerGlnLysAsnLysAlaGlyGlyIleThrLeuProAsp 873
QY 29799 TTCAAACTACATACAGGCTACAGTAACCAACAAACATGCTGCTACCAAAACAGA 29740
Db 874 PheLysLeuTyTrpLysAlaThrValThrLysThrAlaTrpTyTrpTyGlnAsnArg 893
QY 29739 GATATCAACCAATGGAAACAGACAGCCCTCAGAAATAATGCCGATATCTACAACTAT 29680
Db 894 AspIleAspGlnTrpAsnArgThrGluProSerGluIleMetProHisIleTyAsnTy 913
QY 29679 CTGATCTTTTCACAACTGAGAAAAACCAACCAATGGGAAGGATTCCTTATTATAAAA 29620
Db 914 LeuIlePheAspLysProGluLysAsnLysGlnTrpGlyLysAspSerLeuPheAsnLys 933
QY 29619 TGGTGTGTGAGAAAACCTGGCTAGCCCTATGTAGAAAGCTGAACTGGATCCCTCTTACA 29560
Db 934 TrpCysTrpGluAsnTrpLeuAlaIleCysArgLysLeuLysLeuAspProPheLeuThr 953
QY 29559 CCTTATACAAAATTAATTCAAGATGGATTAAAGACTTAATGTTAGACCTAAACCATTA 29500
Db 954 ProTyThrLysIleAsnSerArgTrpIleLysAspLeuAsnValArgProLysThrIle 973
QY 29499 AAATCCCTAGAAAGAAACCCAGCAATACCATTCAGACATAGGCATGGCAGGACTTC 29440
Db 973
```

```
Db 974 LysThrLeuGluLysAsnLeuGlyAsnThrIleGlnAspIleGlyMetGlyLysAspPhe 993
Qy 29439 ATGCTCTAAAACACCAAAAGCAATGGCAACAAAGCCAAAATTCACAAATGGATCTAATT 29380
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 994 MetThrLysThrProLysAlaMetAlaThrLysValLysIleAspArgTrpAspLeuIle 1013
Qy 29379 AACTAAAGAGCTTCTCCACAGCAAAAGAAACACATCAGAGTGTAACAGGCAACTTACA 29320
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1014 LysLeuLysSerPheCysThrAlaLysGluThrThrIleArgValAsnArgGlnProThr 1033
Qy 29319 GAATGGAGAGAAATTTTGGCAACTTACTCATTTGACAAAGGGCTAATATCCAGATCTAC 29260
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1034 LysTrpGluLysIlePheAlaIleTyrSerSerAspLysGlyLeuIleSerArgIleTyr 1053
Qy 29259 AATGAACCTCAACAAATTTTACAGAAAAAAGAAAGAACCCCATCAAAAGAGTGGTGAAG 29200
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1054 AsnGluLeuLysGlnIleTyrLysLysThrAsnAsnProIleLysLysTrpAlaLys 1073
Qy 29199 GATATGAACAGACACTTCTCAAAAGAGGCAATTTATGCAGCCCAAAACACATCAAAAAA 29140
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1074 AspMetAsnArgHisProSerLysGluAspIleTyrAlaAlaLysLysHisMetLysLys 1093
Qy 29139 TACTCATCATCACTGGCCATCAGAGAAATGCAATCAAAACCAACCAATGACATCCTC 29080
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1094 CysSerSerLeuAlaIleArgGluMetGlnIleLysThrThrMetArgTyrHisLeu 1113
Qy 29079 ACACCACTTAGAATGGCAATCATTAAAAAGTCAGGAAACACAGGCTGCTGAGAGATGT 29020
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1114 ThrProValArgMetAlaIleIleLysLysSerGlyAsnAsnArgCysTrpArgGlyCys 1133
Qy 29019 GGAGAAATAGAACACTTTTACACTGTGTGGTGGACTGTGTAAGTGTCAACCATTTGGG 28960
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1134 GlyGluIleGlyThrLeuLeuHisCysTrpTrpAspCysLysLeuValGlnProLeuTrp 1153
Qy 28959 AAGTCAGTGTGGTGAATCTTCACAGGATCTAGAACTAGAAATACCATTTGACCCAGCATC 28900
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1154 LysSerValTrpArgPheLeuArgAspLeuGluLeuGluIleProPheAspProAlaIle 1173
Qy 28899 CCATTACTGGGTATATACCAAGAGGATTATAAATCATGCTGCTATAAAGACACAGACACA 28840
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1174 ProLeuLeuGlyIleTyrProLysAspTyrLysSerCysTyrLysAspThrCysThr 1193
Qy 28839 TGTATGTTTATAGACGACATTCACAAATAGCAAGACTTGGAAACCAACCTAAATGTCCA 28780
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1194 ArgMetPheIleAlaAlaLeuPheThrIleAlaLysThrTrpAsnGlnProLysCysPro 1213
Qy 28779 ACAACGATAGACTGGATTAGAAAAATGTGGCAGATATACACCATGGATGGAATCTAGCAGCC 28720
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1214 ThrMetIleAspTrpIleLysLysMetTrpHisIleTyrThrMetGluTyrTyrAlaAla 1233
Qy 28719 ATAAAAATGATGAGTTCGTCTCTTTGTAGGGACATGGATGAAGCTGGAACACCATCAT 28660
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1234 IleLysAsnAspGluProMetSerProValGlyThrTrpMetLysLeuGluThrIleIle 1253
Qy 28659 CTCAGCAACTATCACAA 28642
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1254 LeuSerLysLeuSerGln 1259
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
RESULT 5
LIN1_NYCCO STANDARD; PRT; 1260 AA.
AC P08548;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-AUG-1988 (Rel. 08, Last annotation update)
DE LINE-1 reverse transcriptase homolog.
OS Nycticebus coucang (Slow loris).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Strepsirhini; Loridae; Nycticebus.
NCBI_TaxID=9470;
RN [1]
RP SEQUENCE.
RX MEDLINE=86230917; PubMed=2423883;
```

```
RA Hattori M., Kuhara S., Takenaka O., Sakaki Y.;
RT "L1 family of repetitive DNA sequences in primates may be derived
RL Nature 321:625-628(1986).
CC -1- MISCELLANEOUS: THIS SEQUENCE WAS CONSTRUCTED FROM AN ALIGNMENT OF
CC SIX SEQUENCES, DETERMINED BY THESE AUTHORS BUT NOT SHOWN,
CC BELONGING TO THE LINE-1 FAMILY.
DR PIR; B25313; GNURL1.
DR HSP; P27695; 1HD7.
DR InterPro; IPR005135; Exo_endo_phos.
DR InterPro; IPR000477; RVTse.
DR Pfam; PF00078; rvt; 1.
DR Pfam; PF03372; Exo_endo_phos; 1.
KW RNA-directed DNA polymerase.
SQ SEQUENCE 1260 AA; 147042 MW; 7A6803DF471F7253 CRC64;

Alignment Scores:
Pred. No.: 1.56e-98 Length: 1260
Score: 1479.50 Matches: 265
Percent Similarity: 78.22% Conservative: 69
Best Local Similarity: 62.06% Mismatches: 92
Query Match: 2.82% Indels: 1
DB: 1

US-10-083-853B-2 (1-29921) x LIN1_NYCCO (1-1260)
Qy 29919 ACTTCTCTCAGCAATTTGGAAGAACTACTTTAAAGTTCATGTGGAACCAAAAGAGCC 29860
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 834 SerTyrPheLysAspLeuGluLysIleLeuHisPheIleTrpAsnGlnLysLysPro 853
Qy 29859 CACATCCCAAGTCAATCTCTAAAGCAAAAGCAAAAGCTGGAGGCATCATCTACCTGCAC 29800
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 854 GlnIleAlaLysThrLeuLeuSerAsnLysAlaGlyGlyIleThrLeuProAsp 873
Qy 29799 TTCAACTACACTACAGGCTACAGTAACCAAAACACATGGTACTGGTACCACAAACAGA 29740
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 874 LeuArgLeuTyrTyrLysSerIleValIleLysThrAlaTrpTyrTrpHisLysAsnArg 893
Qy 29739 GATATCAACCAATGGAACAGACAGAGCCCTCAGAAATAATCGCGCATATCTACCACTAT 29680
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 894 GluValAspValTrpAsnArgIleGluAsnGlnGluMetAspProAlaThrTyrHisTyr 913
Qy 29679 CTGATCTTTGACAAACCTGAGAAACCAAGCAATGGGAAGGATTCCTCTATTATATAA 29620
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 914 LeuIlePheAspLysProIleLysAsnIleGlnTrpGlyLysAspSerLeuPheAsnLys 933
Qy 29619 TGGTGTGAGAAACTGGCTAGCCCTATGTAGAAAGCTGAAAGTGGATCCCTTCTTACA 29560
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 934 TrpCysTrpValAsnTrpLeuAlaIleCysArgArgLeuLysLeuAspProHisLeuSer 953
Qy 29559 CCTTATACAAAATTAATTTCAAGATGGATTAAAGACTTAATGTTAGACCTAAACCATTA 29500
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 954 ProLeuThrLysIleAspSerHisTrpIleLysAspLeuAsnLeuArgHisGluThrIle 973
Qy 29499 AAATCCCTAGAGAAAACCCAGCAATPACCATTCAGACATAGGCATGGCGAAGACTTC 29440
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 974 LysIleLeuGluGluSerAlaGlyLysThrLeuGluGlyIleSerLeuGlyLysTyrPhe 993
Qy 29439 ATGTCTAAAACACCAAAAGCAATGGCAACAAAGCAAAATTCACAAATGGATCTAATT 29380
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 994 MetArgArgTrpProGlnAlaIleGluAlaValSerLysIleHisTyrTrpAspLeuIle 1013
Qy 29379 AAATAAGAGCTTCTCCACAGCAAAAGAAAGCAATCAGAGTGAACAGCACTTACA 29320
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1014 LysLeuLysSerPheCysThrAlaLysAsnIleValSerLysAlaSerArgGlnProSer 1033
Qy 29319 GAATGGAGAGAAATTTTGGCAACTTACTCTATTGACAAAGGGCTAATATCCAGATCTAC 29260
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1034 GluTrpGluLysIlePheAlaGlyTyrThrSerAspLysGlyLeuIleThrArgIleHis 1053
Qy 29259 AATGAACCTCAACAAATTTTACAGAAAAAAGAAAGAACCCCATCAAAAGAGTGGTGAAG 29200
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1054 ArgGluLeuLysHisIleAsnLysLysArgThrArgAspProIleSerGlyTyrAlaArg 1073
```


Alignment Scores:
 Pred. No.: 1.14e-23 Length: 1308
 Score: 443.00 Matches: 198
 Percent Similarity: 38.68% Conservatives: 200
 Best Local Similarity: 19.24% Mismatches: 404
 Query Match: 0.85% Indels: 228
 Db: 1 Gaps: 31

US-10-083-853B-2 (1-29921) x YTX2_XENLA (1-1308)

QY	3455	ATATTAACTTAATAATGTAATGGCTAAATTCCTCCCAATTAATAAGACACAGACTGCGCAAT	3514
Db	5	ILSerThrLeuLeuHisCysArgAsnProPheArgMetPheGlnValLeuSer	24
QY	3515	TGGATAAAGAGTCAAGACCCATCAGTGTCTCTATTACAGAGGCCCATCTCCATCATGA	3574
Db	25	PheLeuArgGlnGlyTyrSerValSerPheLeuGlnGluThrHisThrThrProGlu	44
QY	3575	GACACATAGGCTCAATAAATAAGGATGAGGAGATTTACCAAGTAATGTAAGAACAA	3634
Db	45	LeuGluAlaSerTrpAsnLeuGlu--TrpLysGlyArgValPhePheAsnHisLeuThr	63
QY	3635	AAAAAAAACGACGGGTGCAATCTCTGTATAAACACAGACTTTAAACCAACAAG	3694
Db	64	TrpThrSerCysGlyValThrLeuPheSerAsp-----SerPheGlnProGluVal	81
QY	3695	ATCAAAAGACACAAA-----GAAGGCCATTACATAATGGTAAAGGCATCAATGGAACA	3748
Db	82	LeuSerAlaThrSerValIleProGlyArgLeuLeuHisLeuArgValArgGluSerGly	101
QY	3749	GAGAGCTAATCTCTTAATATACATGACCCCATACAGGA-----GCACCCAGATTC	3802
Db	102	ArgThrTyrAsnLeuMetAsnValTyrAlaProThrThrGlyProGluArgAlaArgPhe	121
QY	3803	ATAAGCAAGTCTTAGAGACCTACAAAGAGACTTTTGACTCCCAACACATAATAGTGGGA	3862
Db	122	PheGluSerLeuSerAlaTyrMetGluThrIleAspSerAspGluAlaLeuIleGly	141
QY	3863	GTCTAAATAATAATAGACACTTTAAACCCCATCTGCAATATTAGGACAGATCAATGAGA	3922
Db	142	-----GlyAspPheAsnTyrThrLeuAspAlaArgAspAsnValPro	156
QY	3923	CAGAAATTAACAGGATATCCAGGAGTTGAAGTCTGAGCTCGGACCAAGCGGACCTAATA	3982
Db	157	LysLysArgAspSerSerGluSerValLeuArgGluLeuIleAlaHisPheSerLeuVal	176
QY	3983	GATATCTACAGAACTCCCAACCCCAATCAACAGATATACACTCTTCTCAGCA-----	4036
Db	177	AspValTrpArgGluGlnAsnProGluThrValAlaPheThrTyrValArgValArgAsp	196
QY	4037	TCATTTACACCTATTATTAATTTAGCATGTAAATTTTAAGTAAACACACTCTCTCAGCAA	4096
Db	197	GlyHisValSerGlnSerArgIleAspArgIleTyrIleSerSerHisLeuMetSerArg	216
QY	4097	TGCAAAAGACAGAAATCCTTAACAAACAGTCTCTCAGACTACAGTGCATCTATTATTGAA	4156
Db	217	AlaGlnSerSerThrIleArgLeuAlaProPheSerAspHisAsnCysValSerLeuArg	236
QY	4157	CTCAGATTAAGAACTCACTCAAAATCACACACTACATGAAACTGAAACACACCTGCTC	4216
Db	237	MetSerIleAlaPro---SerLeuProLysAlaAlaTyrTrpPhePheAsnSerLeu	255
QY	4217	CTGAATGACACTGGGTAAATTAACAAATGAAGGCAAAAATAAAGATGTTCTTTGAAACC	4276
Db	256	LeuGluAspGluGlyPheAlaLysSerValArgAspThrTrpArgGlyTrpArgAlaPhe	275
QY	4277	AATGAGAACAAAGACACAATGTACCAGATCTCTGGGGCATA-----TTTAA	4324
Db	276	GlnAspGluPheAlaThrLeuAsnGln---TrpTrpAspValGlyLysValHisLeuLys	294
QY	4325	GCAGTGTGTAGAGGAAATTTATAGCCTAGATGCTCACAAGAGAAG-----CAG	4375
Db	295	LeuLeuCysGln-----GluTyrThrLysSerValSerGlyGln	307

RESULT 7
 YTX2_XENLA
 ID YTX2_XENLA STANDARD PRT: 1308 AA.
 P14381; AC
 01-JAN-1990 (Rel. 13, Created)
 01-JAN-1990 (Rel. 13, Last sequence update)
 01-APR-1990 (Rel. 14, Last annotation update)
 Transposon TXI hypothetical 149 kDa protein (ORF 2).
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
 OC Xenopodinae; Xenopus.
 NCBI_TaxID=8355;
 [1]
 RN
 SEQUENCE FROM N.A.
 MEDLINE=89384562; PubMed=2550791;
 Garrett J.E., Knutson D.S., Carroll D.;
 "Composite transposable elements in the Xenopus laevis genome.";
 Mol. Cell. Biol. 9:3018-3027(1989).
 [1]
 CC
 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 EMBL; M26915; AAA49976.1; -
 DR PIR; B32494; B32494.
 DR InterPro; IPR005135; Exo_endo_phos.
 DR InterPro; IPR00477; RVTse.
 DR Pfam; PF00078; rvt; 1.
 DR Pfam; PF03372; Exo_endo_phos; 1.
 KW Hypothetical protein; Transposable element.
 SEQUENCE 1308 AA; 149577 MW; C9A9C98CDC169C19 CRC64;


```
ALU7_HUMAN
ID ALU7_HUMAN STANDARD; PRT; 593 AA.
AC P39194;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alu subfamily SQ sequence contamination warning entry.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=95021758; PubMed=7935834;
RA Claverie J.-M.; Makalowski W.;
RT "Alu alert.";
RL Nature 371:752-752(1994).
RN [2]
CONCEPT.
RP MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
ALU FAMILIES CLASSIFICATION.
RP MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation
closely connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
ALU FAMILIES CLASSIFICATION.
RP MEDLINE=91178815; PubMed=1706781;
RA Jurka J.; Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121(1991).
CC -!- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
REPEATS.
CC -!- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CODON, 'XXX', IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -!- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
ACID SEQUENCES.
CC -!- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
BEING REPORTED.
CC -!- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CODING NUCLEOTIDE SEQUENCE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```

```
CC
DR EMBL; U14573; ; NOT_ANNOTATED_CDS.
KW Hypothetical protein.
FT DOMAIN 1 97 FRAME-1.
FT DOMAIN 101 196 FRAME-2.
FT DOMAIN 200 295 FRAME-3.
FT DOMAIN 299 395 FRAME-4.
FT DOMAIN 399 494 FRAME-5.
FT DOMAIN 498 593 FRAME-6.
SQ SEQUENCE 593 AA; 64417 MW; 54A4F50F33A6089F CRC64;
Alignment Scores:
Pred. No.: 1.56e-21 Length: 593
Score: 412.00 Matches: 84
Percent Similarity: 89.80% Conservative: 4
Best Local Similarity: 85.71% Mismatches: 7
Query Match: 0.79% Indels: 3
DB: 1 Gaps: 1
US-10-083-853B-2 (1-29921) x ALU7_HUMAN (1-593)
QY 16853 GGCGGGCGCAGTGGCTCAGCCCTGTAATCTAGCACTTGGGAGCGGAGTGGTGGGA 16794
DB 1 GlyArgAlaArgTrpLeuThrProValIleProAlaLeuTrpGluAlaGluAlaGlyGly 20
QY 16793 TCACCTGAGTTCAGGAGTTCAGAACACAGCGTGGCCCAACATGCGCAAAACCCGCTCTACT 16734
DB 21 SerProGluValArgSerSerArgProAlaTrpProThrTrp***AsnProValSerThr 40
QY 16733 AAAATACAAAATTTAGCCAGCGGTGATGGCAGGACCTGTAAATCCAGCTACTCAGGAG 16674
DB 41 LysAsnThrLysIleSerArgAlaTrpArgAlaProValIleProAlaThrArgGlu 60
QY 16673 GCTGAGCGGGAGAAATCACTTGAACCCGGGAGCGGAGGTCCACTGAGCTGAGTACGCGG 16614
DB 61 AlaGluAlaGlyGluSerLeuGluProGlyArgArgArgLeuGln***AlaGluLeuAla 80
QY 16613 CCATTGCACTCCAGCTGGGTAAAA-----GAAACTCCATCTCAAAAAAAA 16568
DB 81 ProLeuHisSerSerLeuGly-AspLysSerGluThrProSerGlnLysLys 97
RESULT 9
ALU7_HUMAN
ID ALU7_HUMAN STANDARD; PRT; 593 AA.
AC P39194;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alu subfamily SQ sequence contamination warning entry.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=95021758; PubMed=7935834;
RA Claverie J.-M.; Makalowski W.;
RT "Alu alert.";
RL Nature 371:752-752(1994).
RN [2]
CONCEPT.
RP MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
ALU FAMILIES CLASSIFICATION.
RP MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation
closely connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
```


CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
 CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
 CC CODING NUCLEOTIDE SEQUENCE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; U14572; -; NOT_ANNOTATED_CDS.
 KW Hypothetical protein.
 FT DOMAIN 1 97 FRAME-1.
 FT DOMAIN 101 196 FRAME-2.
 FT DOMAIN 200 295 FRAME-3.
 FT DOMAIN 299 395 FRAME-4.
 FT DOMAIN 399 494 FRAME-5.
 FT DOMAIN 498 593 FRAME-6.
 SQ SEQUENCE 593 AA; 64603 MW; 136EF344AACD12A2 CRC64;

Alignment Scores:
 Pred. No.: 4,35e-20 Length: 593
 Score: 392.00 Matches: 77
 Percent Similarity: 82.29% Conservative: 2
 Best Local Similarity: 80.21% Mismatches: 17
 Query Match: 0.75% Indels: 1
 DB: 0 Gaps: 0

US-10-083-853B-2 (1-29921) x ALU6_HUMAN (1-593)

QY 10343 TTTTGTGAGATGGAGTTTCAGTCTTCTCCCGCAGGTAGAGTGCATGCTGATCCC 10402
 DB 299 PhePheLeuArgSerPheAlaLeuValAlaGlnAlaGlnArgAspLeu 318
 QY 10403 GGGTCATGCAACCTCCACCTCCCGGGTCAAGTGTCTCTCCCTCAGCTCCCAAGT 10462
 DB 319 GlySerProGlnProProGlyPheLysArgPheSerCysLeuSerLeuProSer 338
 QY 10463 AGCTGGGACTACAGCAGCTGCCACCGCTGCTTGAACCTCTGACCTCAGGTGATCTGCTGAGAT 10522
 DB 339 SerTrpAspTyrArgHisAlaProArgProAlaAsnPheCysIlePheSerArgAsp 358
 QY 10523 GGGTT-TCCCCACGTTGCCAGGCTGGTCTTGAACCTCTGACCTCAGGTGATCTGCTGCTG 10581
 DB 359 GlyValSerProCysTrpSerGlyTrpSerArgTrpProAspLeuArg***SerAlaArg 378
 QY 10582 CTCAGCCPCCAAATGCTGGGATTACAGGCATGACCCACTGCCGCCG 10629
 DB 379 LeuGlyLeuProLysCysTrpAspTyrArgArgGluProProArgPro 394

RESULT 11
 YOLL_MOUSE
 ID YOLL_MOUSE STANDARD; PRT; 379 AA.
 AC P11260;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 01-JUL-1989 (Rel. 11, Last annotation update)
 DE Hypothetical protein ORF-1137.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP
 RX MEDLINE=87064284; PubMed=3023821;
 RA Loeb D.D., Padgett R.W., Hardies S.C., Shehee W.R., Comer M.B.,
 RA Edgell M.H., Hutchison C.A. III;
 RT "The sequence of a large L1MD element reveals a tandemly repeated 5'
 RT end and several features found in retrotransposons."
 RL Mol. Cell. Biol. 6:168-182(1986).
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; M13002; AAA66023.1; ALT_INIT.
 DR PIR; A24906; QOMSL.
 DR InterPro; IPR004244; Transposase_22.
 DR Pfam; PF02994; transposase_22; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 379 AA; 43788 MW; F4DDE312807644D3 CRC64;

Alignment Scores:
 Pred. No.: 8,74e-20 Length: 379
 Score: 387.00 Matches: 115
 Percent Similarity: 46.74% Conservative: 64
 Best Local Similarity: 30.03% Mismatches: 139
 Query Match: 0.74% Indels: 65
 DB: 1 Gaps: 9

US-10-083-853B-2 (1-29921) x YOLL_MOUSE (1-379)

QY 2301 TCAACACATCAAAAGACC-----AAAGGTAAATAAAAC 2333
 DB 10 ThrValThrThrLysThrAlaSerPheArgAspTyrGlnMetAlaLysGlyLys----- 27
 QY 2334 CAAAAGATGGGAAAAACACAGTGCAGAACACTGAAATTCACAAACACCAACACTCTCT 2393
 DB 28 -----ArgLysAsnProThrAsnArgAsnGlnAsp 37
 QY 2394 TCTCAACCAAGGATCACAACTCTCCGCCACA-----AGGACACAAACACAG 2441
 DB 38 HisSerProSerSerGluArgSerThrProThrProProSerProGlyHisProAsnThr 57
 QY 2442 ATGGAGATGAGTTTGAGGAATTGACAGAGTAGCTTCAGAAAGTGGTGAATACACAAAC 2501
 DB 58 ThrGluAsnLeuAspProAspLeuLysThr-----PheLeuMetMetMetIleGluAsp 75
 QY 2502 -----TCCTCCGAGCTAAAGAGAGATGTTCTAACCCATGCAAGGAAGCT 2546
 DB 76 IleLysLysAspPheHisLysSerLysLysLeuGlnGluSerThrAlaLysGluLeu 95
 QY 2547 AAGAACCTTGAAGAAAGCTTAGATCAATTGCTAACTAGATAATACAGTGTAGAGAGAAC 2606
 DB 96 GlnAlaLeuLysGluLysGlnGluAsnThrAlaLysGlnValMetGluMetAsnLysThr 115
 QY 2607 ATAAATGACCTG-----ATG 2621
 DB 116 IleLeuGluLeuLysGlyGluValAspThrIleLysLysThrGlnSerGluAlaThrLeu 135
 QY 2622 GAGCTGAAAAACGCAAGACAAAGACTTCATGAAGCATACACAGCTTCAATAGCCAAATC 2681
 DB 136 GluIleGluThrLeuGlyLysArgSerGlyThrIleAspAlaSerIleSerAsnArgIle 155
 QY 2682 GATCAACGACAGAAAGAGATATCAGTGATTGAAGATCAATTAATAAAGAAAGTGCAGAA 2741
 DB 156 GlnGluMetGluGluArgIleSerGlyAlaGluAspSerIleGluAsn----- 171
 QY 2742 GACAAGATTACAGAAAAAGAGTGAAGAAAGAAC-----AAACAA 2780
 DB 172 -----IleAspThrThrValLysGluAsnThrLysCysLysArgIleLeuThrGln 188
 QY 2781 AGCTCCCAAGAAATTATGGGACTATGTGAAAGACCAAACTACATTTGATTTGGTGTCCC 2840
 DB 189 AsnIleGlnValIleGlnAspThrMetArgArgProAsnLeuArgIleIleGlyLeuAsp 208
 QY 2841 CAAGATGATGGGAGAAATGGAATCACTGAGTGGAAACACTCTTCAGGGGTATTATCCAGGAG 2900
 DB 209 GluAsnGluAspPheGlnLeuLysGlyProAlaAsnIlePheAsnLysIleIleGluGlu 228

QY 2901 AATTCCCC---ATCTATCAGGCGAGCCCAACATTCAGAAATATGGAGAACAC 2957
 D 229 AsnProAsnLysLysGluMetProMetile--IleGlnGluAlaThrArgThrP 248
 QY 2958 CATAAAGTACTCTCGAGAGACATCCCAACACATTAATCTTCAGATTCACCAAGG 3017
 D 248 roAsnArgLeuAspGlnLysArgAsnSerArgHisIleIleIleIleIleIleIle 268
 QY 3018 TTGAATAGAGAAAAATGTTAAGGCGAGCCAGAGAGAGAGTTGGTTACCCACAAAG 3077
 D 268 laLeuAsnLysAspArgIleLeuLysAlaValArgGluLysGlyGlnValThrTyrLysG 288
 QY 3078 GGAAGCCATCAGACTAACAGCGGATCCCGGAGAAACCTCAAGCCAGAGAGAGAGT 3137
 D 288 lyArgProIleArgIleThrProAspPheSerProGluThrMetLysAlaArgAlat 308
 QY 3138 GAGGCCAATATTCACATCTTAAAGAAATATTTTCAACCCAGAAATTCATATCCAG 3197
 D 308 rpThrAspValIleGlnThrLeuArgGluHisLysCysGlnProArgLeuLysTyrProA 328
 QY 3198 CCAACCAAGCTCTCTAAGTGAAGGAGAGAAATAAATCTCTACAGAGAACCAATGCTGA 3257
 D 328 laLysLeuSerIleThrIleAspGlyGluThrLysValPheHisAspLysThrLysPheT 348
 QY 3258 CAGATTTTGTACACACAGCGCTCTTCAAGAGAGCTCTGAGAGAGAGAGAGAGAGT 3317
 D 348 hrGln-TyrLeuSerThrAsnProAlaLeuGlnArgIleIleThrGluLysLysGlnTyr 367
 QY 3318 AAA 3320
 D 368 Lys 368

RESULT 12

ALU6_HUMAN STANDARD; PRT; 593 AA.

AC P39193; 1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Alu subfamily SP sequence contamination warning entry.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95021758; PubMed=7935834;
 RA Claverie J.-M.; Makalowski W.;
 RT "Alu alert."; Genomics 12:838-841(1992).
 RL Nature 371:752-752(1994).
 RN [2]
 RP CONCEPT.
 RX MEDLINE=92241891; PubMed=1572661;
 RA Claverie J.-M.;
 RT "Identifying coding exons by similarity search: alu-derived and other
 RT potentially misleading protein sequences."; Genomics 12:838-841(1992).
 RL [3]
 RP ALU FAMILIES CLASSIFICATION.
 RX MEDLINE=88333009; PubMed=3138422;
 RA Quentin Y.;
 RT "The Alu family developed through successive waves of fixation
 RT closely connected with primate lineage history."; J. Mol. Evol. 27:194-202(1988).
 RL [4]
 RP ALU FAMILIES CLASSIFICATION.
 RX MEDLINE=91178815; PubMed=1706781;
 RA Jurka J.; Milosavljevic A.;
 RT "Reconstruction and analysis of human Alu genes."; J. Mol. Evol. 32:105-121(1991).
 RL [5]
 CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
 CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
 CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX

CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
 CC REPEATS.
 CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
 CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
 CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
 CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
 CC ACID SEQUENCES.
 CC -1- CAUTION: ALU REPEAT SEQUENCES ARE INTERSPERSED IN HUMAN AND
 CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
 CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
 CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
 CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
 CC LIGATED WITH ALU-DERIVED SEQUENCES IN ANY ORIENTATION. ALTHOUGH ALU
 CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
 CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS.
 CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
 CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
 CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
 CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
 CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
 CC BEING REPORTED.
 CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
 CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
 CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
 CC CODING NUCLEOTIDE SEQUENCE.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U14572; ; NOT_ANNOTATED_CDS.
 DR Hypothetical protein.
 KW DOMAIN 1 97 FRAME-1.
 FT DOMAIN 101 196 FRAME-2.
 FT DOMAIN 200 295 FRAME-3.
 FT DOMAIN 299 395 FRAME-4.
 FT DOMAIN 399 494 FRAME-5.
 FT DOMAIN 498 593 FRAME-6.
 SQ SEQUENCE 593 AA; 64603 MW; 136EF344AACD12A2 CRC64;
 Alignment Scores:
 Seq. No.: 1 39e-19 Length: 593
 Score: 385.00 Matches: 80
 Percent Similarity: 87.76% Conservatives: 6
 Best Local Similarity: 81.63% Mismatches: 9
 Query Match: 0.73% Indels: 3
 DB: 1 Gaps: 1
 US-10-083-853B-2 (1-29921) x ALU6_HUMAN (1-593)
 QY 16853 GGCCGGGCGCAGTGGCTCAGCGCTTAATCTGACTTTGGAGGCCGAGGTGGTGA 16794
 D 1 GlyArgAlaArgTrpLeuThrProValIleProAlaLeuTrpGluAlaGlyGly 20
 QY 16793 TCACCTGAGGTGAGAGTTCAAGACAGCGTGGCCACATGGCAAAACCCGCTCTACT 16734
 D 21 SerProGluValGlySerSerArgProAla***ProThrTrpArgAsnProValSerThr 40
 QY 16733 AAAATACAAAAATTAGCCAGGCGTGTATGCGACGACCTTAATCCAGCTACTCAGGAG 16674
 D 41 LysAsnThrLysIleSerArgAlaTrpTrpArgMetProValIleProAlaThrArgGlu 60
 QY 16673 GCTGAGCGGGAGAGTCACTTGAACCCGGAGCGGAGGTTCGAGTCAGTCAGATCGCG 16614
 D 61 AlaGluAlaGlyGluSerLeuGluProGlyArgArgArgLeuArg***AlaGluIleAla 80
 QY 16613 CCATTCCACTCCAGCTGGGTAAAA-----GAACTCCATCTCAAAAAAAA 16568
 D 81 ProLeuHisSerSerLeuGly-AsnLysSerGluThrProSerGlnLysLys 97

```
RESULT 13
ALU8_HUMAN
ID AC P39195;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Alu subfamily SX sequence contamination warning entry.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95021758; PubMed=7935834;
RA Claverie J.-M.; Makalowski W.;
RT "Alu alert.";
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation
closely connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J.; Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121(1991).
CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
REPEATS.
CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CODON, 'xxx', IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
ACID SEQUENCES.
CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
BEING REPORTED.
CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CODING NUCLEOTIDE SEQUENCE.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
```

```
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
CC
CC EMBL: U14574; -, NOT_ANNOTATED_CDS.
CC
DR KW Hypothetical protein.
FT DOMAIN 1 96 FRAME-1.
FT DOMAIN 100 195 FRAME-2.
FT DOMAIN 199 294 FRAME-3.
FT DOMAIN 298 393 FRAME-4.
FT DOMAIN 397 492 FRAME-5.
FT DOMAIN 496 591 FRAME-6.
SQ SEQUENCE 591 AA; 64395 MW; AC8154AD8A6B280 CRC64;

Alignment Scores:
Pred. No.: 2,49e-19 Length: 591
Score: 381.50 Matches: 79
Percent Similarity: 85.57% Conservative: 4
Best Local Similarity: 81.44% Mismatches: 10
Query Match: 0.73% Indels: 4
DB: Gaps: 1

US-10-083-853B-2 (1-29921) x ALU8_HUMAN (1-591)
QY 16854 AGCCGGGCGCAGTGGCTCAGCCTGTAATCCAGCACTTTGGAGCGCGAGTGGTGG 16795
Db 4 ArgProGlyAlaValAlaHisAlaCysAsnProSerThrLeuGlyGly-Arg----- 20
QY 16794 ATCACTGAGTGCAGGAGTTCACAGACAGCGGTGCCCAACATGCCAAACCCGCTCTAC 16735
Db 21 -SerProGluValArgSerArgProAlaTrpThrTrp***AsnProValSerTh 40
QY 16734 TAAAAATACAAAATATAGCCAGCGCTGATGCAGGACCTGTATCCAGCTACTCAGGA 16675
Db 40 rLysAsnThrLysIleSerArgAlaTrpArgAlaProValIleProAlaThrArgG 60
QY 16674 GGCTAGCGCGGAGAAATCACTTGAACCCGGAGCGGAGGTGCAGTGCAGTGCAGATCGC 16615
Db 60 uAlaGluAlaGlyGluSerLeuGluProGlyArgArgLeuGln***AlaGluIleAl 80
QY 16614 GCATTCCTCCAGCTCCAGCTGGTAAAGAAACTCCATCTCAAAAAA 16566
Db 80 aProLeuHisSerSerLeuGlyAspArgAlaArgLeuArgLeuLysLys 96

RESULT 14
ALU5_HUMAN
ID ALU5_HUMAN STANDARD; PRT; 585 AA.
AC P39192;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alu subfamily SC sequence contamination warning entry.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95021758; PubMed=7935834;
RA Claverie J.-M.; Makalowski W.;
RT "Alu alert.";
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation
```


CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC BEING REPORTED.
CC -!- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CC CODING NUCLEOTIDE SEQUENCE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U14571; -; NOT_ANNOTATED_CDS.
KW Hypothetical protein.
FT DOMAIN 1 95 FRAME-1.
FT DOMAIN 99 193 FRAME-2.
FT DOMAIN 197 291 FRAME-3.
FT DOMAIN 235 389 FRAME-4.
FT DOMAIN 393 487 FRAME-5.
FT DOMAIN 491 585 FRAME-6.
SQ SEQUENCE 585 AA; 63957 MW; 46E8C4F493650A7 CRC64;

Alignment Scores:

Pred. No.: 6e-17 Length: 585
Score: 348.50 Matches: 75
Percent Similarity: 78.35% Conservative: 1
Best Local Similarity: 77.32% Mismatches: 18
Query Match: 0.67% Indels: 3
DB: .1 Gaps: 1

US-10-083-853B-2 (1-29921) x ALU5_HUMAN (1-585)

QY 16569 TTTTTCAGATGGAGTTCTTTT---ACCCAGGCTGGAGTCAATGGCGGATCTCAGC 16625
Db 393 PhePhe***AspGlyValSerLeuCysArgGlnAlaGlyValGlnTrpArgAspLeuGly 412
QY 16626 TCACGTGCAACCTCCGCCCTCCCGGTTCAAGTGAATCTCCCGCCTCAGCCTCTCTGAGTAGC 16685
Db 413 SerLeuGlnProProProGlyPheLysArgPheSerCysLeuSerLeuProSerSer 432
QY 16686 TGGGATTACAGTGCCTGCCATCAGCCTGCGCTAATTTTGTATTTTAGTAGACACGGG 16745
Db 433 TrpAspTyrArgAlaProArgProAlaAsnPheCysIlePheSerArgAspGly 452
QY 16746 GTTTTGGCCATGTGGCCACGCTGTCTTGAACCTCCTGACCTCAGGTGATCCACCCACCTC 16805
Db 453 ValSerProCysTrpProGlyTrpSerArgSerLeuAspLeu--ValIleArgProProA 472
QY 16806 GGCCTCCCAAGATGCTAGATTACAGGCTGAGCCACTGCGCCCGGC 16852
Db 472 rgProProLysValLeuGlyLeuGlnAla***AlaThrAlaProGly 487

Search completed: July 4, 2003, 20:08:09
Job time : 462.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 4, 2003, 20:07:18 ; Search time 429 Seconds
(without alignments)
13409.968 Million cell updates/sec

US-10-083-853B-2

Sequence: 1 grtatgtgaagaagctca.....caattctgtgaagaagtaa 29921

Scoring parameters: BLOSUM62
Gapop 10.0, Xgapext 0.5
Gapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh
-Q/cq2_1/USPTO.spool/US10083853/runat_03072003_093612.8103/app_query.fasta.1.30087
-DB=PIR73 -QFMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=BLOSUM62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USPR=US10083853.ecgn.1.1.723.ernat_03072003_093612.8103 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: PIR73
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5336	10.2	1280	2 B34087	hypothetical prote
2	5195	9.9	1275	2 S65824	reverse transcript
3	5169	9.9	1275	2 I38588	reverse transcript
4	5139	9.8	1259	4 GNHUL1	retrovirus-related
5	5108	9.8	1275	2 B28096	line-1 protein ORF
6	3658.5	7.0	1260	4 GNLR1	retrovirus-related
7	3552.5	6.8	1281	1 GNMSLL	retrovirus-related
8	2880.5	5.5	712	2 S23650	retrovirus-related
9	2257	4.3	1280	2 B34087	hypothetical prote
10	2229	4.3	1275	2 S65824	reverse transcript
11	2215	4.2	1275	2 I38588	reverse transcript
12	2196	4.2	1275	2 B28096	line-1 protein ORF
13	2193	4.2	712	2 S23650	retrovirus-related
14	2176.5	4.2	562	2 JU0033	hypothetical L1 pr

C	15	2145	4.1	1259	4 GNHUL1	retrovirus-related
	16	1932	3.7	685	2 S16783	probable RNA-direc
	17	1632.5	3.1	556	2 S21347	hypothetical prote
C	18	1479.5	2.8	1260	4 GNLR1	retrovirus-related
	19	1466.5	2.8	513	2 S21976	probable RNA-direc
C	20	1450.5	2.8	1281	1 GNMSLL	retrovirus-related
	21	1190.5	2.3	500	2 S16788	probable reverse t
	22	1119.5	2.1	1258	2 T14853	reverse transcript
C	23	1117	2.1	562	2 JU0033	hypothetical L1 pr
	24	1104.5	2.1	1258	2 T14855	reverse transcript
	25	1090.5	2.1	338	2 I38587	retrovirus-related
C	26	995.5	1.9	275	2 S21348	probable pol polyp
	27	972	1.9	335	4 A26718	retrovirus-related
	28	969.5	1.9	340	2 PH0217	reverse transcript
C	29	962.5	1.8	392	2 S57662	hypothetical prote
	30	850	1.6	275	2 S21348	probable pol polyp
	31	826.5	1.6	392	2 S57662	hypothetical prote
C	32	746	1.4	513	2 S21976	probable RNA-direc
	33	704	1.3	143	2 E41925	hypothetical prote
C	34	636	1.2	143	2 E41925	hypothetical prote
	35	485.5	0.9	138	2 I59767	lamin-like protein
	36	463.5	0.9	202	2 S21346	probable pol polyp
	37	443	0.8	1308	2 B32494	transposable eleme
	38	439.5	0.8	190	2 B25556	hypothetical prote
	39	422.5	0.8	1152	2 S20106	hypothetical prote
	40	416	0.8	351	2 S21345	retrovirus-related
	41	415.5	0.8	1231	2 C84716	hypothetical prote
C	42	409	0.8	370	2 JC7591	spinal cord-derive
	43	409	0.8	762	2 T47440	hypothetical prote
	44	399.5	0.8	872	2 H86435	protein F17F8.5 [1
	45	399	0.8	1344	2 H84557	hypothetical prote

ALIGNMENTS

RESULT 1

B34087
hypothetical protein (L1H 3' region) - human
C:Species: Homo sapiens (man)
C>Date: 30-Mar-1990 #sequence_revision 30-Mar-1990 #text_change 30-Sep-1993
C:Accession: B34087
R:Scott, A.F.; Schmeckpeper, B.J.; Abdelrazik, M.; Conney, C.T.; O'Hara, B.; Rossiter, Genomics 1, 113-125, 1987
A:Title: Origin of the human L1 elements: proposed progenitor genes deduced from a co A:Reference number: A34087; MUID:88085185; PMID:3692483
A:Accession: B34087
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-1280 <SCO>
C:Superfamily: pol polyprotein

Alignment Scores:
Pred. No.: 0 Length: 1280
Score: 5336.00 Matches: 1084
Percent Similarity: 79.94% Conservative: 68
Best Local Similarity: 75.23% Mismatches: 123
Query Match: 10.20% Indels: 173
DB: 2 Gaps: 7

US-10-083-853B-2 (1-29921) x B34087 (1-1280)

QY	3413	CCAGCTAGTGTCTCATATGCGAGCATCAATTCACACATATATATTAACCTTAATGTA	3472
DB	1	ProAlaAsnIleileMetThrGlySerAsnSerHisIleThrIleLeuThrLeuAsnVal	20
QY	3473	ATFGGCTAAATTCCTCCCAATTAAGACACAGACTGGCAATTTGGATTAAGAGTCAAGAC	3532
DB	21	AsnGlyLeuAsnAlaProIleLysArgHisArgLeuAlaAsnTrpIleLysSerGlnAsp	40
QY	3533	CCATCAGTGTGCTGTATTTCAGGAGGCCCATCTCACATGAAGACACACATAGGCTCAA	3592
DB	41	ProSerValCysCysIleGlnGluThrHisLeuMetCysArgAspThrHisArgLeuLys	60

```
QY 3593 ATAAAGGGATGGAGAGATTATACCAAGTAAATGGGAAACCAAAAAAGCAGGGTT 3652
Db 61 IleLysGlyTrpArgLysIleTyrGlnAlaAsnGlyLysGln---LysLysAlaGlyVal 79
QY 3653 GCAATCCTAGTCTCTGATATAAAGAGACTTTAAACCAACAAGATCAAAAGACAGAA 3712
Db 80 AlaIleuValSerAspLysThrAspPheLysProThrLysIleLysArgAspLysGlu 99
QY 3713 GGCCATTACATATAGGAGCATCAATGGAGCAAGAGAGCTAACTATCTTAATATA 3772
Db 100 GlyHisTyrIleMetValLysGlySerIleGlnGlnGluGluLeuThrIleLeuAsnIle 119
QY 3773 CATGCACCAATACAGAGACCCAGATTCAATAAGCAAGTCTTACAGACCTACAAAGA 3832
Db 120 TyrAlaProAsnThrGlyAlaProArgPheIleLysGlnValLeuSerAspLeuGlnArg 139
QY 3833 GACTTTGACTCCACACATAATAGTGGGAGTCTAAATAATAATAGACACTTTACACC 3892
Db 140 AspLeuAspSerHisThrIleIleMetGly-----AspPheAsnThr 153
QY 3893 CCACTGCCAATATTAGGCAGATCAATGACAGAGAAATTAACAAGGATATCCAGAGTTG 3952
Db 154 ProLeuSerThrLeuAspArgSerThrArgGlnLysValAsnLysAspIleGlnGluLeu 173
QY 3953 AACTGAGCTCTGGACCAAGCGGACCTAATAGATATCTACAGAACTCCCAACCAATCA 4012
Db 174 AsnSerAlaLeuHisGlnAlaAspLeuIleAspIleTyrArgThrLeuHisProLysSer 193
QY 4013 ACAGATATACACTCTCTCAGCATCACATACACTATTTTAAATTCACCATGAATT 4072
Db 194 ThrGluTyrPhePheSerAlaProHisHisThrTyrSerLysIleAspHisIleLeu 213
QY 4073 TTAAGTAAACACTCTCTCAGCAATGCAAAAGAACAGAAATCTTAACAACAGTCTCTCA 4132
Db 214 GlySerAlaLeuLeuSerLysCysLysArgThrGluIleIleThrAsnCysLeuSer 233
QY 4133 GACTACAGTGCATCTATTAGACTCAGAAATTAAGAACTCACTCAAAATCACACA 4192
Db 234 AspHisSerAlaIleLysLeuGluLeuArgIleLysLysLeuThrGlnAsnHisSerThr 253
QY 4193 ACATGAAACTGAACAACCTCTCTCTGATGACTACTGGTAAATAACAATAAGAGCA 4252
Db 254 ThrTrpLysLeuAsnLeuLeuAsnAspTyrTrpValHisAsnGluMetLysAla 273
QY 4253 AAATAAAGATGTCTTTGAAACCAATGAGAACAGACACAATGTACCAAGTCTCTGG 4312
Db 274 GluIleLysAsnPhePheGluThrAsnGluAsnLysAspThrTyrGlnAsnLeuTrp 293
QY 4313 GGCATATTAAAGCAGTGTGTAGAGGAAATTTATAGCACTAGATGCCTACAGAGAAG 4372
Db 294 AspThrPheLysAlaValCysArgGlyLysPheIleAlaLeuAsnAlaHisLysArgLys 313
QY 4373 CAGGAAATCTAAATAGACACCTTAACATCACAAATTAAGAAGTACAGAGAAGAG 4432
Db 314 GlnGluArgSerLysIleAspThrLeuThrSerGlnLeuLysGluLeuGlyGlnGlu 333
QY 4433 CAACAATAATCAAAAGCTAGCAGAGACAGAAATAACTAAGATCAGACAGAACTGAAG 4492
Db 334 GlnThrHisSerLysAlaSerArgGlnGluIleThrLysIleArgAlaGluLeuLys 353
QY 4493 GAGATAGACACAAAAAGCCCTTCAATAAATCAATCAATCCAGAGCTGGTTTTTGA 4552
Db 354 GluIleGluThrGlnLysThrLeuGlnLys-IleAsnGluSerArgSerTrpPheGlu 373
QY 4553 AAAGATCAGCAAAAT---AGACCCTAGACAGACTAATAAGAGAAAGAGAGAGAA 4608
Db 373 uLysIleAsnLysIleAspArgProLeuAlaArgLeuIleLysLysGlyGluLysAs 393
QY 4609 TCAAGAGATGCAATAAAAAATGATAAAGGGGATATCACACCGATCCACAGAAATACA 4668
Db 393 nGlnIleAspSerIleLysAsnAspLysGlyAspIleThrAspProThrGluIleGlu 413
QY 4669 AACTATTATCAGAGAAATATTATAAACACCTCTATGCAAAATAAAGTATAGAGA 4728
```

623	snLysLeuGlyIleAspGlyThrTyrLeuLysIleIleArgAlaIleTyrAspLysProT	643
5805	CAGCCAAATATCATACTGAATGGGCAAAAACCTGGAAGCATTCCTCTTTGAGAACCTGGCACAA	5864
643	hrAlaAsnIleIleLeuAsnGlyGlnLysLeuGluAlaPheProLeuLysThrGlyThra	663
5865	GACAAGATGCCCTCTCTCACCACTCCTATTCGAAGATACACTATTGGAAGTTCTCGCCAGGG	5924
663	rgGlnGlyCysProLeuSerProLeuLeuPheAsnIleValLeuGluValLeuAlaArga	683
5925	CAATCAGGCAATAGAAAGAAATAAAGGTATTCAAATAGAAAGAGGAGGTATCATATTGT	5984
683	laIleArgGlnGluLysGluIleLysGlyIleGlnLeuGlyLysGluIleValLysLeuS	703
5985	CTCTGTTTGCAGATGACATGTTGTATATTAGAAAACCCCATCTCTCAGGCCAAAAAC	6044
703	erLeuPheAlaAspAspMetIleValTyrLeuGluAsnProIleValSerAlaGlnAsnL	723
6045	TCCTTAAGCTGATAAGCAACTTCAGCAAAAGTCTCAGGACACAAAATAATGTGCAAAAAT	6104
723	euLeuLysLeuIleSerAsnPheSerLysValSerGlyTyrLysIleAsnValGlnLysS	743
6105	CACAAGCATTCATTAGCCCAATAATAGACAAACAGAGAGCCAAAATCATGACTCT	6164
743	erGlnAlaPheLeuTyrThrAsnAsnArgGlnThrGluSerGlnIleMetSerGluLeuP	763
6165	CATTCAAAATGCTACAAAGAGAATAAAATACCTAGGAATACAACTTACAAGGACACAGT	6224
763	roPheThrIleAlaSerLysArgIleLysTyrLeuGlyIleGlnLeuThrArgAspValL	783
6225	AGGAATCTTCAAGGAGAAGTACAACCACTGAAGGAAATPAGAGAGGACACAAAACA	6284
783	ysAspLeuPheLysGluAsnTyrLysProLeuLeuAsnGluIleLysGluAspThrAsnL	803
6285	AATGGAAAACATTCCTCATCTCAGATAGTACAGATCTAGAAATCAT-----GAAAAATGCCATAC	6337
803	ysTrpLysAsnIleProCysSerTrpValGlyArgIleAsnIleValLysMetAlaIleL	823
6338	TGCCAAAGTAAATATPAGATTCAGTGCTACCCCCCATCAAGCTACCATTCACATTCCTTCA	6397
823	euProLysValIleTyrArgPheAsnAlaIleProIleLysLeuProMetThrPhePheT	843
6398	CAGAATTGGAAAAACAACCTTTAAATTTCTATGTGAACCAAAAAAGAGCCACAGAGCC	6457
843	hrGluLeuGluLysThrThrLeuLysPheIleTrpAsnGln-LysArgAlaHisIleAla	862
6458	AAGACATCTTAGCAAAAAGAACAAAGCTGGAGGTATCATGTCTACTGACTTAAAACTA	6517
863	LysSerIleLeuSerGlnLysAsnLysAlaGlyIleThrLeuProAspPheLysLeu	882
6518	TACTATAAGCTACAGTAACCAAACTGCATGTCTGTTGGTACCACAAACAGATATATAGAC	6577
883	TyrTyrLysAlaThrValThrLysThrAlaTrpTyrTrpTyrGlnAsnArgAspIleAsp	902
6578	CAATGGAAACAGAACAGAGACCTCAGAAATTAC-ACTGCAATCTACATCCCATCTGATCTTT	6636
903	GlnTrpAsnArgThrGluProSerGluIleMetProHisIleTyrAsnTyrLeuIlePhe	922
6637	GACAAACCTGACAAAACAGCAATGGAAAAAGGATTCCTCTATTTAATAATGGTGTGG	6696
923	AspLysProAspLysAsnLysGlnTrpGlyLysAspSerLeuPheAsnLysTrpCysTrp	942
6697	AAAAATGGCTAGCCATATGCAGAAAGCTCAAACTGGATCCCTCTCTTACACCTTATACA	6756
943	GluAsnTrpLeuAlaIleCysArgLysLeuLysLeuAspProPheLeuThrProTyrThr	962
6757	AAAGTTAACTCAAGATGAATTAAGACTTAAATATAGACATATAAACCATAAAACCCCA-	6815
963	LysIleAsnSerArgTrpIleLysAspLeuAsnValArgProLysThrIleLysThrLeu	982
6816	GAGAAAACCTAGCAATACCATTCAGGATATGGACATGGCAAGAGACTTCATGACTAAA	6875
983	GluGluAsnLeuGlyAsnThrIleGlnAspIleGlnMetGlyLysAspPheMetThrLys	1002

Qy	6876	ACACCAAAAGCAATGGCAACAAAGCCAAAAATAGACAAGTGGGATCTGATTAACACTATAG	69335
Db	1003	ThrProLysAlaMetAlaThrLysAlaLysIleAspLysTrpAspLeuLleLysLeuLys	1022
Qy	6936	AGCTTCGCACAGCAAAAAAAAACTGTCATCAGAGTGAACAAGCAACCTACAGAATGGGA	6995
Db	1023	SerPheCysThrAla-LysGluThrThrIleLeuValAsnArgLInProThrGluTrpGI	1042
Qy	6996	GAAATTTTTCGAATCTATCGATCTGCACAAAGGCTAATATCCAGAGATCTACGAAGAAC	7055
Db	1042	uLysIlePheAlaIleIyrSerSerAspLysGlyLeuIleSerArgIleIyrAsnGluLe	1062
Qy	7056	TAAACAATTTACAGAAAAA-----AACACCCCGCTCAAAATATGGCCAAGATATGAG	7111
Db	1062	uLysGlnIleIyrLysLysLysThrAsnAsnProIleLysLysTrpAlaLysAspMetAs	1082
Qy	7112	CAGACACTTCTCAAAAGACACATTTATGCAGCGACAAACAACATATGAAAAAACCTCATC	7171
Db	1082	naRGHisPheSerLysGluAspIleIyrAlaAlaAsnLysHisMetLysLysCysSerSe	1102
Qy	7172	ATCATTTGGTCGTTAGAGAAATGCAAAACAAAAACACAGTCACATCATCTCATGCTAGT	7231
Db	1102	rSerLeuAlaIleArgGluMetGlnIleLysThrThrMetArgTyrHisLeuThrProVa	1122
Qy	7232	TAGAAATGGTGATCACTAAAAAGTCAGAAACAACAATACTGTCGAGAGGATGTGGAGAAAT	7291
Db	1122	lArgMetAlaIleLleLysLysSerGlyAsnAsnArgCysTrpArgGlyCysGlyGluIl	1142
Qy	7292	AGGAACACTTTTCCACTGTTGGTGGGAATGTAATAGTTCAACACCTGTTGGGAAGACAGT	7351
Db	1142	eGlyThrLeuLeuHisCysTrpTrpAspCysLysLeuValGlnProLeuTrpLysThrVa	1162
Qy	7352	GTGAGAGATCCTTAAGGATCTAGAACCCAGAAATATCATTTGACCCAGCAATCCCATTACT	7411
Db	1162	lTrpArgPheLeuLysAspLeuGlnLeuGluIleProPheAspProAlaIleProLeuLe	1182
Qy	7412	GAGTATATCCCAAGGAATATAAATCATCTCTATTATAAAGACACATGCACACATATGTT	7471
Db	1182	uGlyIleTrpProLysAspTrpLysSerCysCysTyrLysAspThrCysThrArgMetPh	1202
Qy	7472	TATTGCAGCACTGATCAATAGCAAAAGACTTGGAAACCCCAAAATGTCCATCAGTGTAT	7531
Db	1202	eIleAlaAlaLeuPheThrIleAlaLysThrTrpAsnGlnProLysCysProSerMetIl	1222
Qy	7532	AGACTGGATAAAGAAAACATGGCCACATATACACCATGAATACTATGCCACCCATAAAAG	7591
Db	1222	eAspTrpIleLysLysMetTrpHisIleTyrThrMetGluTyrTyraAlaIleLysAs	1242
Qy	7592	-GATGAGTTCATGTCCTTCGACAGATATCGATGAAGCTGGAACCATCATCTCTCAGCAA	7650
Db	1242	naSpGluPheMetSerPheValGlyThrTrpMetLysLeuGluThrIlelleLeuSerLy	1262
Qy	7651	ACTAACACAAGAACAGAAAAACCAACACCATGTTCTCACTTGTAAAGTGGGAT	7705
Db	1262	sLeuSerGlnGlyGlnLysThrLysHisArgMetPheSerLeuIleGlyValasn	1280

RESULT 2

S65824 reverse transcriptase homolog - human transposon L1.1
C:Species: Homo sapiens (man)
C:Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 2
C:Accession: S65824
R:Dombroski, B. A.
submitted to the EMBL Data Library, January 1992
A:Description: Isolation of an active human transposable element.
A:Reference number: S65823
A:Accession: S65824
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1275 <DOM>
A:Cross-references: EMBL:M80340; NID:g339767; PIDN:AAA51622.1; PIR:Superfamily: pol polyprotein

Alignment Scores:

Pred. No.:	0	Length:	1275
Score:	5195.00	Matches:	1057
Percent Similarity:	78.97%	Conservative:	77
Best Local Similarity:	73.61%	Mismatches:	136
Query Match:	9.93%	Indels:	173
DB:	2	Gaps:	7

US-10-083-853B-2 (1-29921) x S65824 (1-1275)

3428	ATGCGAGGATCAAAATTCACACATAAATAATATTAACCTTAATGTAAATGGCTAAATTC	3487
1	MetThrGlySerAsnSerHisIleThrIleLeuThrLeuAsnIleAsnGlyLeuAsnSer	20
3488	CCAAATTAAGACACAGACTGGCAAAATTGGATAAAGATCAAGACCCATCAGTGTGCTGT	3547
21	AlaIleLysArgHisArgLeuAlaSerTrpIleLysSerGlnAspProSerValCysCys	40
3548	ATTGAGGAGCCCATCTCACATGAAGAAGACACACATAGGCTCAAAATAAAGGATGGAGG	3607
41	IleGlnGluThrHisLeuThrCysArgAspThrHisArgLeuLysIleLysGlyTrpArg	60
3608	AAGATTACCAAGTAATGGAAAAACAAAAAACCAGGGGTGGCAATCCTAGTCTCT	3667
61	LysIleTyrGlnAlaAsnGlyLysGln---	79
3668	GATFAAACAGACTTTAAACCAACAAAGATCAAAAGACACAAAGGCCATTACATAATG	3727
80	AspLysThrAspPheLysProThrLysIleLysArgAspLysGluGlyHisTyrIleMet	99
3728	GTAAGGCATCAATGGAACAAGAGAGCTTAACCTATCTTAATATACATGACCCCAATACA	3787
100	ValLysGlySerIleGlnGlnGluLeuThrIleLeuAsnIleTyrAlaProAsnThr	119
3788	GGAGCACCAGATTCAATGAACAAGTCTTATAGAGACCTACAAAGAGACTTTGACTCCAC	3847
120	GlyAlaProArgPheIleLysGlnValLeuSerAspLeuGlnArgAspLeuAspSerHis	139
3848	ACAATAATAGTGGAGTCTAAATAATAAATAGACACTTTAACACCCCACTGCCAAATATTA	3907
140	ThrIleLeuMetGly-----AspPheAsnThrProLeuSerThrLeu	153
3908	GGCAGATCAATGAGACAGAAAAATTACAAGGATATCCAGGAGTTGAAGTCTGGAC	3967
154	AspArgSerThrArgGlnLysValAsnLysAspThrGlnGluLeuAsnSerAlaLeuHis	173
3968	CAAGCGGACTAATAGATATCTACAGAACTCCCAACCCCAAAATCAACAGAAATATACACT	4027
174	GlnAlaAspLeuIleAspIleTyrArgThrLeuHisProLysSerThrGluTyrThrPhe	193
4028	TTCTCAGATCATCATACACCTATTTTAAATTTGACCATGTAAATTTAAGTAAACACTC	4087
194	PheSerAlaProHisIleThrTyrSerLysIleAspHisIleValGlySerLysAlaLeu	213
4088	CTCAGCAATGCAAAAGACAGAAATCCTPAACAAACAGTCTCTCAGACTACAGTGAATC	4147
214	LeuSerLysCysLysArgThrGluIleIleThrAsnTyrLeuSerAspHisSerAlaIle	233
4148	TATTTAGAAGTCAAGAAATTAAGAACTCACTCAAAATACACAACTACATGGAACTGAAC	4207
234	LysLeuGluLeuArgIleLysAsnLeuThrGlnSerArgSerThrThrTrpLysLeuAsn	253
4208	AACTCTGCTGTAATGACTACTGGTGAATAATAACAAATAAGAGCGAAATAAAGATGTTTC	4267
254	AsnLeuLeuLeuAsnAspTyrTrpValHisAsnGluMetLysAlaGluIleLysMetPhe	273
4268	TTTGAACCAATCAGAACAAAGACACAATGTACCAGAACTCTGCGGGCATATTTAAAGCA	4327
274	PheGluThrAsnGlnAsnLysAspThrThrTyrGlnAsnLeuTrpAspAlaPheLysAla	293
4328	GTGTGTAGGGGAAATTTTATAGCACTAGATGCCGTACAAGAGAAAGCAGGAAATATCTAAA	4387

Db	294	valCysargGlyLysPheIleAlaLeuAsnAlaTyrlLysargLysGlnGluArgSerLys	311
Qy	4388	ATAGACACCTTAACATCACATTAATAAAGAACTAGAGAAGAAAGACGACAAATTCAAAA	4447
Db	314	IleAspThrLeuThrSerGlnLeuLysGluLeuGluLysGlnGluInThrHisSerLys	333
Qy	4448	GCTAGCAGACACAAGAATAACTAAGATCAGAGCAGAACTGGAAGGAGATAGACAGACAA	4507
Db	334	AlaSerArgargGlnGluIleThrLysIleargAlaGluLeuLysGluIleGluThrGln	353
Qy	4508	AAAGCCCTTCAATAAATCAATGAATCCAGGAGCTGCTTTTTCAAAAGATCAGCAAAAT	4567
Db	354	LysThrLeuGlnLys-IleasnGluSerArgSerTrpPhePheGluargIleasnLysIi	373
Qy	4568	----AGACCACTAGACAGACTAATAAAGAGAAGAAAGAGAAATCAAAGAGATGCAAT	4623
Db	373	eAsPargProLeuSerargLeuIleLysLysLysArgGluLysasnGlnIleAspThrIi	393
Qy	4624	AAAAAATGATAAGGGGATATCACACCAGATCCACAGAAATACAAACTATTATCAGAGA	4683
Db	393	eLysAsnAspLysGlyAspIleThrThrAspProThrGluIleGlnThrThrIleArgG	413
Qy	4684	ATATTATAACACCTCTATGCAATAAATACTAGAAATCTAGAGAAGAAATGGATAAATTCCT	4743
Db	413	uTyrlLysHisLeuTyrlAlaasnLysLeuGluasnLeuGluMetAspThrPheLe	433
Qy	4744	GGACACATATGTAGCCTGTATGGACCTTGGGGACAGACAAAGGGGTCAATGCAGAA	4803
Db	433	uAsp-----	434
Qy	4804	ATAAAGACAAAGACAAAGAGATATGTTGGAAGTAGGGGTCAAGGGGGCAACTTGCCTCT	4863
Db	434	-----	434
Qy	4864	AATGGACAAGGGCCCTGAGCTTTACACCAACCCCTCTGTATTATTAGGCAAAAGAGATAGC	4923
Db	434	-----	434
Qy	4924	GAGAGGGTGAGTTGGAAGAAGAGGTGAGCTGTTAGGTCACAGATAGCCCTGCAAGACTGC	4983
Db	434	-----	434
Qy	4984	ATTCCTCAACAATAGGCTCTAGATGTCCCAAGTAGATAACCTCAAGGAGCCAGTGCCAGG	5043
Db	434	-----	434
Qy	5044	GAGTGATGGCCCTCAGCAAAACCTTCTAGGCGAGCGACAGAGAAGTAAGTTTGCCCACTTCT	5103
Db	434	-----	434
Qy	5104	GTATTACGATAAACAGCTTTTGCTTTTGATCAAGTAGCCTCCAGTGGAATGCTCAGTTGG	5163
Db	434	-----	434
Qy	5164	TCATGATCCCTTTTGGCCTTCTTGGCTCCCAAAACACATACACCTCTCAAGACTAAACCA	5223
Db	435	-----ThrTyrlThrLeuProargLeuasnG	443
Qy	5224	GGAGAAGTCAAAATCCCTGAAATATACAGTACAAAGTTCTAAATTTGAACAGCAATTTGA	5283
Db	443	nGluGluValGluSerLeuasnArgProIleThrGlySerGluIleValAlaIleIleAs	463
Qy	5284	TAGCTACCAACCAAAAAAGTCCAGGACAGCGGATTCACAGCCAAATTCACACAGAG	5343
Db	463	nSerLeuProThrLysLysSerProGlyProAspGlyPheThrAlaGluPheTyrlGlnAr	483
Qy	5344	GTACAAGAAGAAGCTGCTACTATTCTCTTCTGAAACTATTCCAAAAAATGAAA---AATGG	5400
Db	483	gTyrlMetGluGluLeuValProPheLeuLeuLysLeuPheGlnSerIleGluLysGlu-u	503
Qy	5401	GAATCCTCCCTAACTATTTTACAGGCCAGCATCATCTGTATACCAAAAACCTAGCAGTG	5460
Db	503	lvIleLeuProAsnSerPheTyrlGluAlaSerIleIleLeuIleProLysProGlyVarq	523

QY	5461	ACACACAAAAAGAGGAAATTTTCAGGCCCATATCCCTGATGAAACATTTGATGCAAAATCC	5520	QY	6533	GTAAACAAAAGCTGCTACTGGTACCAACAAAACAGATATATAGACAACTGGACAGAACA	6592
Db	523	sphrThrLysLysGluAsnPheArgProIleSerLeuMetAsnIleAspAlaLysIleL	543	Db	883	ValThrLysThrAlaTrpTrpTyrGlnAsnArgAspIleAspGlnTrpAsnArgThr	902
QY	5521	TCAATAAATACTGGCAACCAAAATCCAGCAGACACATCAAAAAGCTTATCTACCATGATC	5580	QY	6593	GAGACCTCAGAAATTTAC-ACTGCAATCTACATCCATCTGATCTTTTGACAAAACCTGACAA	6651
Db	543	euAsnLysIleLeuAlaAsnArgIleGlnGlnHisIleLysLysLeuIleHisAspG	563	Db	903	GluProSerGluIleMetProHisIleTyrAsnTyrLeuIlePheAspLysProGluLys	922
QY	5581	AAGTTGGGCTCATCCCTGGGATGCAAGGCTGCTCAAAATATGCAAAATCAATAATAGTAG	5640	QY	6652	AACAAGCAATGGAAAAAGGATTCCCTTATTATAATGTTGGTGGAAAAAATGGCTGAGC	6711
Db	563	lnValGlyPheIleProGlyMetGlnGlyTrpPheAsnIleArgLysSerIleAsnValI	583	Db	923	AsnLysGlnTrpGlyLysAspSerLeuPheAsnLysTrpCysTrpGluAsnTrpLeuAla	942
QY	5641	GCCATCACATAAACAGAACCATGATGACAAAAACACATGATTTATCTCAATAGATGAGAA	5700	QY	6712	ATATCGAGAAAGCTGAAACTGGATCCCTTCCCTTACACCTTATACAAAAGTTAACTCAAG	6771
Db	583	leGlnHisIleAsnArgAlaAsnAspLysAsnHisMetIleIleSerIleAspAlaGluL	603	Db	943	IleCysArgLysLeuLysLeuAspProPheLeuThrProTyrThrLysIleAsnSerArg	962
QY	5701	AGGCCCTTGTCAAAATTCACAGCCCTTCATGCTTAAATTTCTAGTAAACTAGGTATCG	5760	QY	6772	TGAATTAAGACTTAAATATAGACATATAAACCATAAACCCTAATAAACCCA-GAAGAAACCTGAGC	6830
Db	603	ysAlaPheAspLysIleGlnGlnProPheMetLeuLysThrLeuAsnLysLeuGlyIleA	623	Db	963	TrpIleLysAspLeuAsnValLysProLysTrpIleLysThrLeuGluGluAsnLeuGly	982
QY	5761	ATGGAAATGATCTCAAAATAATAAGAGCTATTATTATAC-AAACCCACAGCCAATATCATAC	5819	QY	6831	AATACCATTCAGGATATGGACATGGCAAGACTTTCATGACTTAAACACCAACCAAGCAATG	6890
Db	623	spGlyThrPheLysIleIleArgAlaIleTyrAspLysProThrAlaAsnIleIleL	643	Db	983	IleThrIleGlnAspIleGlyValGlyLysAspPheMetSerLysThrProLysAlaMet	1002
QY	5820	TGAATGGCAAAACTGGAAGCATTCCTCTTGGAGAACTGGCACAGCAAGGATGCCCTC	5879	QY	6891	GCACAAAAGCCAAATAGACAAGTGGGATCTGATTAAACTATAGAGCTTCTGCACAGCA	6950
Db	643	euAsnGlyGlnLysLeuGluAlaPheProLeuLysThrGlyThrArgGlnGlyCysProL	663	Db	1003	AlaThrLysAspLysIleAspLysTrpAspLeuIleLysLeuLysSerPheCysThrAla	1022
QY	5880	TCTCACCCTCTANTTCAGATACTATTGGAAGTTCTGGCCAGGGCAATCAGGCAATAGA	5939	QY	6951	AAAAAACTGTCATCAGAGTGAACAGCAACTACAGAACTGGGAGAAATTTTGCAT	7010
Db	663	euserProLeuLeuPheAsnIleValLeuGluValLeuAlaArgAlaIleArgGlnGluL	683	Db	1023	-LysGluThrThrIleArgValAsnArgGlnProThrTrpGluLysIlePheAlaTh	1042
QY	5940	AAGAAATAAGGCTATTCAAAATAGAAAGAGGAAGTATATTGCTCTGTTGGAGATG	5999	QY	7011	CTATCGATCTGACAAAGGCTAATATCCAGAGATCTACGAAGAACTTAAACAAATTTACAA	7070
Db	683	ysGluIleLysGlyIleGlnLeuGlyLysGluGluValLysLeuSerLeuPheAlaAspA	703	Db	1042	rTyrSerSerAspLysGlyLeuIleSerArgIleTyrAsnGluLeuLysGlnIleTyrL	1062
QY	6000	ACATGTTGTATATTTAGAAAACCCCATCTGCTCAGGCCCAAAACTCTTAAAGCTGATAA	6059	QY	7071	GAAAAA---AACAACCCCGTCAAAATATGCGCAAGGATATGAGCAGACACTTCTCAA	7126
Db	703	spMetIleValTyrLeuGluAsnProIleValSerAlaGlnAsnLeuLysLeuLys	723	Db	1062	siLysLysThrAsnAsnProIleLysLysTrpAlaLysAspMetAsnArgHisPheSerL	1082
QY	6060	GCAACTTCAGAACTCTCAGCACACAATCAATCAATGTCGCAAAATCACAGCATTTCTAT	6119	QY	7127	AGACACATTTATGAGCCCAACACACATATGAAAAAACCTCATCATCATCTGCTGTTAG	7186
Db	723	erAsnPheSerLysValSerGlyTyrLysIleAsnValGlnLysSerGlnAlaPheLeuT	743	Db	1082	sgLyspIleTyrAlaAlaLysLysHisMetLysLysCysSerSerSerLeuAlaIleAr	1102
QY	6120	ACGCCAATAATAGACAAACAGAGCCCAATCATGAGTGAATCTCATTCACAAATGCTA	6179	QY	7187	AGAAATGCAAAAACAAACCCACAGTGCATACATCTCATGCTAGTTAGTAGTGGTATCAC	7246
Db	743	yrThrAsnAsnArgGlnThrGluSerGlnIleMetClyGluLeuProPheValIleAla	763	Db	1102	gGluMetGlnIleLysThrThrMetArgTyrHisLeuThrProValArgMetAlaIleL	1122
QY	6180	CAAAGAGATAAATACCTAGGAATACAACTTACAAGGGACACGCTAGGAACCTCTCAAG	6239	QY	7247	TAAAGTCAGGAAACACAAATGCTGGAGAGATGTGGAGAAATAGGAACTTTTCCA	7306
Db	763	erLysArgIleLysTyrLeuGlyIleGlnLeuThrArgAspValLysAspLeuPheLysG	783	Db	1122	eLysLysSerGlyAsnAsnArgCysTrpArgGlyCysGlyGlyIleGlyThrLeuLeuH	1142
QY	6240	AGAACTACAAACCTGATCAGGAATAAAGAGAGGACACAAACAATGGAAAAACATTC	6299	QY	7307	CTGTTGGTGGGAATTAATTTAGTTCAACCTGTTGGAGACAGCTGCTGAGATTCTCTAA	7366
Db	783	luAsnTyrLysProLeuLeuLysGluIleLysGluAspThrAsnLysTrpLysAsnIleP	803	Db	1142	sCysTrpTrpAspCysLysLeuValGlnProLeuTrpLysSerValTrpArgPheLeuAr	1162
QY	6300	CATGCTCAGATAGTAAGAATCAT-----GAAAAATGCCATCTGCCCCCAAGTAAAT	6352	QY	7367	GGATCTAGAACCCAGAAATATCATTTGACCCCAATCCCATCTACTGAGTATATACCCAAA	7426
Db	803	roCysSerTrpValGlyArgIleAsnIleValLysMetAlaIleLeuProLysValIleT	823	Db	1162	gasPLeuGluLeuGluIleProPheAspProAlaIleProLeuLeuGlyIleTyrProAs	1182
QY	6353	ATAGATTAGTCTACCCCATCAAGCTACCATTTGACTTTCTTCACAGAAATTTGAAAAA	6412	QY	7427	GGAATATAAATCATTTATTAATAAGACACATGACACATATGTTTATTTGACGACTGAT	7486
Db	823	yrArgPheAsnAlaIleProIleLysLeuProMetThrPhePheThrGluLeuGluLysT	843	Db	1182	ngLutyrLysSerCysCysTyrLysAspThrCysThrArgMetPheIleAlaAlaLeuPh	1202
QY	6413	CAACTTTAAATTCATATGGACCAAAAAAGAGCCACAGACCAAGCAATCTTAAGC	6472	QY	7487	CACAATAGCAAGACTTGGAAACCAACCCCAAAATGTCCATCAGTCATAGACTGATTAAGA	7546
Db	843	hrThrLeuLysPheIleTrpAsnGln-LysArgAlaArgIleAlaLysSerIleLeuSer	862	Db	1202	ethrIleAlaLysThrTrpAsnGlnProLysCysProThrMetIleAspTrpIleLysL	1222
QY	6473	AAAAAGACAAAGCTGGAGGTATCATGCTTACCTGACTTAAACTATATTAAGCTACA	6532	QY	7547	AACATGGCACATATACACCATCAAAATCTATGACGCAATAAAAG-GATGAGTTCATGTC	7605
Db	863	GlnLysAsnLysAlaGlyIleThrLeuProAspPheLysLeuTyrTrpLysAlaThr	882	Db	1222	smetTrpHisIleTyrThrMetGluTyrTrpAlaAlaIleLysAsnAspGluPheIleSe	1242
Db				QY	7606	CTTTCAGAGATATGGATGAAGCTGAAACCACTCATCTCTCAGCAAACTAACACAAACA	7665

QY	5104	GTATTCAAGATAAACAGCTTTGCTGTTTGATGTCAGTAGCTCCAGTCCGAATGCTGAGTTGG	5163
DB	434	-----	434
QY	5164	TCATGATGTCCTTTGGCCTCCAAAAACACATACACCCCTCTCAAGACTAAACCA	5223
DB	435	-----ThrryrrhrleuProarglleuysGL	443
QY	5224	GGAGAAGTCAAAATCCCTGTAATACACAGTAACAAGTCTTAAATTTGAAGCAGCAATTTGA	5283
DB	443	nGluValGluSerLeuAsnGlyProIleThrGlySerGluIleValAlaIleIleAs	463
QY	5284	TAGCTTACCACCAAAAAAGTCCAGCAGCAGCGGATTCACAGCCAAATCTTACCAGAG	5343
DB	463	nSerLeuProThrLysSerProGlyProAspGlyPheThrAlaGluPheTyrrGlnAr	483
QY	5344	GTCAAAAGAGAAGCTGTACTATTCTCTTCTGAAACATTCCTCAAAAAATGAAA	5400
DB	483	gTyrrLysGluGluLeuValProPheLeuLeuLysLeuPheGlnSerIleGluLysGlu-G	503
QY	5401	GAATCTCTCCCTACTCATTTTACGAGGCCAGCATCATCTGTATACCAAAAACCTAGCAGTG	5460
DB	503	lyIleLeuProAsnSerPheTyrrGluAlaSerIleIleLeuIleProLysProGlyArgA	523
QY	5461	ACACAACAAAAAGGAAATTTGAGGCCCATATCCCTGATGAACATTTGATGAAATCC	5520
DB	523	spThrThrLysLysGluAsnPheArgProIleSerLeuMetAsnIleAspAlaLysIleL	543
QY	5521	TCATAAAATCTGGCAACCAATCCAGCAGCACATCAAAAAGCTTATCTACCATGATC	5580
DB	543	euAsnLysIleLeuAlaAsnArgIleGlnHisIleLysLysLeuIleHisHisaspG	563
QY	5581	AAGTTGGCGTATCCCTGGGATGCAAGCGTGGTTCAAATATGCAATCAATAAATGTAG	5640
DB	563	lnValGlyPheIleProGlyMetGlnGlyTrpPheAsnIleIleArgLysSerIleAsnValI	583
QY	5641	GCCATCATATAACAGAACCAATGACAAAAACACATGATTCTCAATAGATGCCAGAAA	5700
DB	583	leGlnHisIleAsnArgAlaLysAspLysAsnHisMetIleIleSerIleAspAlaGluL	603
QY	5701	AGGCCTTTGTCAAATTCACAGCCCTTCATGCTAAAAATTTCTCAGTAAACTAGGTATCG	5760
DB	603	ysAlaPheAspLysIleGlnProPheMetLeuLysThrLeuAsnLysLeuGlyIleA	623
QY	5761	ATGGAATGTATCTCAAAATTAATAGAGCTATTATATAC-AAACCCACAGCCAATCATATC	5819
DB	623	spGlyThrTyrrPheLysIleIleArgAlaIleTyrrAspLysProThrAlaAsnIleIleL	643
QY	5820	TGAATGGCAAAAACCTGGAGCATTCCTTTGAGAACTGGCACAAAGCAAGGATGCCCTC	5879
DB	643	euAsnGlyGlnLysLeuGluAlaPheProLeuLysThrGlyThrArgGlnGlyCysProL	663
QY	5880	TCTCACCACTCTTATTCAGATACTATTGGAAGTCTTGCCAGGCGCAATCAGGCAATAGA	5939
DB	663	euSerProLeuLeuPheAsnIleValLeuGluValLeuAlaArgAlaIleArgGlnGluL	683
QY	5940	AAGAAATAAGGGTATTCAATAGAACAGAGGAGTCAATTTGTCTCTCTTTTGCAGATG	5999
DB	683	ysGluIleLysGlyIleGlnLeuGlyLysGluGluValLysLeuSerLeuPheAlaAspA	703
QY	6000	ACATGTTTGTATTTAGAAAAACCCCATCTCTCAGGCCCAAAAACCTCTTAAAGCTGATAA	6059
DB	703	spMetIleValTyrrLeuGluAsnProIleValSerAlaGlnAsnLeuLeuLysLeuLys	723
QY	6060	GCAACTTCAGAAAGTCTCAGGACACAAAATCAATGTGCAAAAATCACAAGCATTTCTTAT	6119
DB	723	erAsnPheSerLysValSerGlyTyrrLysIleAsnValGlnLysSerGlnAlaPheLeuT	743
QY	6120	ACGCCATATATAGACAACAGAGCAACCAATCATGATGTAACCTCTCATTCACAATTCGCTA	6179
DB	743	yrThrAsnAsnArgGlnThrGluSerGlnIleMetGlyGluLeuPheProPheThrIleAlas	763

6180	QY	CAAGAGAGATAAATACCTAGGATACAACTTACAGGGACACGCTAGCACTCTTCAAGG	62339
763	Db	erLysArgILeLysTyrLeuGlyIleGlnLeuThrArgAspValLysAspLeuPheLysG	783
6240	QY	AGAACTACAAACCACTCATCAAGCAATAAGAGAGGACACAAACAAATGGAAAAACATTC	62999
783	Db	luasnTyrLysProLeuLeuLysGluLeuLysGluAspThrAsnLysTrpLysAsnLleP	803
6300	QY	CATGCTCCACAGATAGTAAGATCAT-----GAAATGCCATACTGCCCCAAAGTAAATT	6352
803	Db	roCysSerTrpValGlyArgIleAsnIleValLysMetAlaIleLeuProLysValIleT	823
6353	QY	ATGATTCACTGCTACCCCATCAAGCTACATTGACTTCTTCACACAGATTGGAAAAA	6412
823	Db	yrArgPheAsnAlaIleProIleLysLeuProMetThrPhePheThrGluLeuGluLysT	843
6413	QY	CAACTTTAAATTTTCATATGGAACCAAAAAAGAGCCACAGAGCAAGACAACTTTAAGC	6472
843	Db	hrThrLeuLysPheIleTrpAsnGln-LysArgAlaArgIleAlaLysSerIleLeuSer	862
6473	QY	AAAAAGACAAAGCTGGAGGTATCATCTGCTACCTGACTTAAACTATATCTATAAGGCTACA	6532
863	Db	GlnLysAsnLysAlaGlyGlyIleThrLeuProTyrPheLysLeuTyrTyrLysAlaThr	882
6533	QY	GTAACCAAACTGCATGGTACTGTGTACCAAAACAGATATATAGACCAATGGACAGAAC	6592
883	Db	ValThrLysThrAlaTrpTyrTrpTyrGlnAsnArgAspIleAspGlnTrpAsnArgThr	902
6593	QY	GAGACCTCAGAAATTAC-ACTGCAATCTACATCCATCTGATCTTTGACAACCTGACAAA	6651
903	Db	GluproSerGluIleMetProHisIleTyrAsnTyrLeuIlePheAspLysProGluLys	922
6652	QY	AACAGCAATGGAAAAAGATTCCCTATTAAATAAATGGTGTGGAAAAACTGGCTAGCC	6711
923	Db	AsnLysGlnTrpGlyLysAspSerLeuPheAsnLysTrpCysTrpGluAsnTrpLeuAla	942
6712	QY	ATATGCAGAAAGCTGAACCTGGATCCCTCTTACACCTTATACAAAGTTTAACCTCAAG	6771
943	Db	IleCysArgLysLeuLysLeuAspLeuPheLeuThrProTyrThrLysIleAsnSerArg	962
6772	QY	TGAATTAAGACTTAATATATAAGACATAAAACCAATAAAACCCA-GAAGAAAACTTAGC	6830
963	Db	TrpIleLysAspLeuAsnValLysProLysThrIleLysThrLeuGluGluAsnLeuGly	982
6831	QY	AATACCATTCAGGATATGGACATGGCAAGACTTCATGACTTAAACACCAACCAAGCAATG	6890
983	Db	IleThrIleGlnAspIleGlyValGlyLysAspPheMetSerLysThrProLysAlaMet	1002
6891	QY	GCACAAAGCCAAATAGACAACTGGGATCTGATTAACTATAGACCTTCTGCACAGCA	6950
1003	Db	AlaThrLysAspLysIleAspLysTrpAspLeuIleLysLeuLysSerPheCysThrAla	1022
6951	QY	AAAAAAACTCTCATCAGATGACACAGCAACCTACAGAATGGGAGAAAATTTTGCAT	7010
1023	Db	-LysGluThrThrIleArgValAsnArgGlnProThrThrTrpGluLysIlePheAlaThr	1042
7011	QY	CTATCGATTCACAAAGGCTAATATCCAGAGATCTACGAAGACTTAAACAAAATTTACAA	7070
1042	Db	rTy-SerSerAspLysGlyLeuIleSerArgIleTyrAsnGluLeuLysGlnIleTyrLys	1062
7071	QY	GAAGAAAA----AACACCCCGTCAAAATATGGCAAGGATATAGCAGACACTTCTCAAA	7126
1062	Db	sLysLysThrAsnAsnProIleLysLysTrpAlaLysAspMetAsnArgHisPheSerLys	1082
7127	QY	AGAGACATTTATCGACCCCAACAACATATGCAAAAAAACCTCATCATTCATTGGCTGTAG	7186
1082	Db	sGluAspIleTyrAlaAlaLysLysHisMetLysLysCysSerSerSerLeuAlaIleAr	1102
7187	QY	AGAAATGCAAAACAAAACACAGTGCATACCATCTCATGCTAGTTAGATTGATGTCAC	7246
1102	Db	gLuMetGlnIleLysThrThrMetArgTyrHisLeuThrProValArgMetAlaIleI	1122
7247	QY	TAAAAAGTCAGGAACAACAAATGCTGGAGAGGATGTGGAGAAATAGGAACACTTTTCCA	7306

Db	1122	eLYsYsSerGlyAsnAsnArgCysTrpArgGlyCysGlyGluileGlyThrLeuLeuHi	1142
Qy	7307	CTGTGGTGGGAATGTAAATTAAGTCAACCATTTGTGAAGACAGTGTGGAGATTCTCTAA	7366
Db	1142	sCYsTrpTrpAspCysLYsLeuValGlnProLeuTrpLYsSerValTrpArgPheLeuAr	1162
Qy	7367	GGATCTAGAACCAAGAATATCATTTGCCACGAGCAATCCCATTAAGTATATACCAAA	7426
Db	1162	gAspLeuGluLeuGluileProPheAspProAlaileProLeuLeuGlyileTrpGlg	1182
Qy	7427	GGAAATAAATCATTTCTATTATAAGACACATCCACACATATGTTTATTGCAGCACTGAT	7486
Db	1182	uAspTrpLYsSerCysCYsTrpLYsAspThrCysThrArgMetPheileAlaLeuPh	1202
Qy	7487	CACAANTAGCAAGACATTGGAACCAACCCCAATGTCCATCAGTCATAGACTGGATAAGAA	7546
Db	1202	eThrileAlaLYsThrTrpAsnGlnProLYsCysProThrMetileAspTrpLeuLYsly	1222
Qy	7547	AAACATGGCACATATACACCATGAATACTATGCAGGCATAAAAAG-GATGAGTTCAATGC	7605
Db	1222	smetTrpHisileTrpThrMetGluTrpTyrAlaAlaileLYsAsnAspGluPheilese	1242
Qy	7606	CTTTGGCAGATATGGATGAAGTGGAAACCATCATCTCAGCAAACTAACCAAGAA	7665
Db	1242	rPheValGlyThrTrpMetLYsLeuGluThrileileLeuSerLYsLeuSerGlnGluG	1262
Qy	7666	GAANAACCAACACCATGTTCTCACTTGTAACTGGGAGT	7705
Db	1262	nLYsTrpLYsHisArgilePheSerLeuileGlyGlyVasn	1275

RESULT 4

GNRHU1
retrovirus-related reverse transcriptase pseudogene - human
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1988 #sequence_revision 04-Jan-1996 #text_change 14-May-1999
C:Accession: A25313
R:Hattori, M.; Kuhara, S.; Takenaka, O.; Sakaki, Y.
Nature 321, 625-628, 1986
A:Title: Ll family of repetitive DNA sequences in primates may be derived from a sequence
A:Reference number: A93381; MUID:86230917; PMID:2423883
A:Accession: A25313
A:Status: conceptual translation of pseudogene
A:Molecule type: DNA
A:Residues: 1-1259 <RAW>
A:Note: this sequence was constructed from an alignment of published and unpublished seq
C:Keywords: reverse transcriptase; pseudogene

Alignment Scores:		
Pred. No.:	0	Length: 1359
Score:	5139.00	Matches: 1050
Percent Similarity:	79.00%	Conservative: 71
Best Local Similarity:	74.00%	Mismatches: 132
Query Match:	9.82%	Indels: 173
DB:	4	Gaps: 7

US-10-083-853B-2 (1-29921) x GNHUL1 (1-1259)

[illegible]


```
Db 1004 LysValLysIleAspArgTrpAspLeuIleLysLeuLysSerPheCysThrAla-LysGI 1023
QY AACTGTCATCAGAGTGAACAAGCAACCTACAGAAATGGAGAAATTTTGGCAATCATCG 7016
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1023 uThrThrIleArgValAsnArgGlnProThrLysTrpGluLysIlePheAlaIleYrse 1043
QY ATCTGACAAGGCTAATATCCAGAGATCTACGAGAACTTAAACAATAATTTACAGAAAAA 7076
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1043 rSerAspLysGlyLeuIleSerArgIleYrAsnGluLeuLysGlnIleYrLysLysly 1063
QY 7077 ----RACACCCCGTCAAAATATGGCAAGGATATGACGACACATTTCTCAAAAGAAGA 7132
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1063 sThrAsnAsnProIleLysTrpAlaLysAspMetAsnArgHisProSerLysGluAs 1083
QY 7133 CATTATGACGCCACAACAACATATGAAAAAACCCTCATCATCTTGGTGTAGAGAAAT 7192
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1083 pIleTyAlaAlaLysHisMetLysLysCysSerSerLeuAlaIleArgGluMe 1103
QY 7193 GCAAAACAACACAGTGCATACCATCTCATGCTAGTAGTAATGGTGATCACTAAAAA 7252
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1103 tGlnIleLysThrMetArgTyHisLeuThrProValArgMetAlaIleIleLysLy 1123
QY 7253 GTGAGAAACAACAATGCTGGAGAGATCTGGAGAAATAGGAAACACTTTCCACTGTG 7312
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1123 sSerGlyAsnAsnArgCysTrpArgGlyCysGlyGluIleGlyThrLeuLeuHisCysTr 1143
QY 7313 GTGGGAATGTAATATTAGTTCAACATTGTGGAAGACAGTGTGGAGATTCCTTAAGGATCT 7372
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1143 pTrpAspCysLysLeuValGlnProLeuTrpLysSerValTrpArgPheLeuArgAspLe 1163
QY 7373 AGAACCAAAATATCATTTGACCCCAATCCCATCTAGTATATATATATATATATATATAT 7432
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1163 uGluLeuGluIleProPheAspProAlaIleProLeuLeuGlyIleTyProLysAspTy 1183
QY 7433 TAAATCATCTATTATAAGACACATGCACACATATGTTTATGTCAGCACTGATCACAAT 7492
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1183 rLysSerCysCysTyLysAspThrCysThrArgMetPheIleAlaAlaLeuPheThrIl 1203
QY 7493 AGCAAGACTTGGAAACCAACCCCAATGTCCATCAGTATAGATGATGATGATGATGATG 7552
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1203 eAlaLysThrTrpAsnGlnProLysCysProThrMetIleAspTrpIleLysLysMetTr 1223
QY 7553 GCACATATACACCATGAATATCTATCCAGCCATATAAAG-GATGAGTTTCATGTCCTTGC 7611
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1223 pHisIleTyThrMetGluTyTrAlaAlaIleLysAsnAspGluProMetSerProVa 1243
QY 7612 AGAGATATGGATGAAGCTGGAACCATCTCTCAGCAAACTAACACAA 7660
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1243 lGlyThrTrpMetLysLeuGluThrIleIleLeuSerLysLeuSerGln 1259

RESULT 5
B28096
line-1 protein ORF2 - human
C:Species: Homo sapiens (man)
C:Date: 03-Nov-1988 #sequence_revision 03-Nov-1988 #text_change 30-Sep-1993
C:Accession: B28096
R:Skowronski, J.; Fanning, T.G.; Singer, M.F.
Mol. Cell. Biol. 8, 1385-1397, 1988
A:Title: Unit-length line-1 transcripts in human teratocarcinoma cells.
A:Reference number: A28096; MUID:88246405; PMID:2454389
A:Accession: B28096
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1275 <SKO>
C:Superfamily: pol polyprotein

Alignment Scores:
Pred. No.: 0 Length: 1275
Score: 5108.00 Matches: 1045
Percent Similarity: 78.64% Conservative: 85
Best Local Similarity: 72.72% Mismatches: 140
Query Match: 9.76% Indels: 175
DB: 2 Gaps: 7
```

```
US-10-083-853B-2 (1-29921) x B28096 (1-1275)
QY 3428 ATGGCAGGATCAAAATTCACACATAATAATTAACCTTAATTAATGGCTAAATTC 3487
Db 1 MetThrGlySerAsnSerHisIleThrIleLeuThrLeuAsnValAsnGlyLeuAsnSer 20
QY 3488 CCAATTAAGACACAGACTGGCAATTTGGATAAAGAGTCAAGACCCATCAGTGTCTGT 3547
Db 21 ProIleLysArgHisArgLeuAlaSerTrpIleLysSerGlnAspProSerValCysCys 40
QY 3548 ATTCAGAGGCCATCTCACATGAAAGACACACATAGGCTCAAAATAAAGGATGAGG 3607
Db 41 IleGlnGluThrHisLeuMetCysArgAspThrHisArgLeuLysIleLysGlyTrpArg 60
QY 3608 AGATTTCACCAAGTAAATGGAACAAAAAAGAGGGGTTTCCAATCCTAGTCTCT 3667
Db 61 LysIleTyTrpGlnAlaAsnGlyLysGln---LysLysAlaGlyValAlaIleLeuValSer 79
QY 3668 GATAAACACAGACTTTAAACCAACAAGATCAAAAGACACAAAGAGCCCATACATAATG 3727
Db 80 AspLysThrAspPheLysProThrLysIleLysArgAspLysGluGlyHisTyMetMet 99
QY 3728 GTAAGGCAATCAATGGAACAAGAGCTAATCTTAAATATATACATGCACCATATACA 3787
Db 100 ValLysGlySerIleGlnGluLeuThrIleLeuAsnMetTyAlaProAsnThr 119
QY 3788 GGAGCACCCAGATTCTAAAGCAAGTTCTTAGAGACTTACAAAGACACTTTGACTCCAC 3847
Db 120 GlyAlaProArgPheIleLysGlnValLeuSerAspLeuGlnArgAspLeuAspSerHis 139
QY 3848 ACAATAATAGTGGGAGTCTTAAATAATAATAGACACTTTTAACACCCCACTGCC-AAATAT 3906
Db 140 ThrLeuIleMetGly-----AspPheAsnThrProLeuSerThrLeu 153
QY 3907 AGCAGATCAATGACACAGAAAAATTAACAAGATATCCAGAGTTCAGACTGAGCTCTGGA 3966
Db 154 ArgGlnIleAsnGluThrGluSerGlnGlnGlyTyTr-GlnGluLeuAsnSerAlaLeuH1 173
QY 3967 CCAAGGGACCTTAATAGATATCTACAGAATCTCCCAACCCCAATCAACAGATATACACT 4026
Db 173 sGlnAlaAspLeuIleAspIleTyArgThrLeuHisProLysSerThrGluTyThrPh 193
QY 4027 CTCTCAGCATCACATTACACCTATTTTAAATTTGACCATGTAATTTTAAAGTAAACACT 4086
Db 193 ePheSerAlaProHisHisSerTySerLysIleAspHisIleLeuGlySerGluAlaLe 213
QY 4087 CCTCAGCAATGCAAAAGACAGAAATCTTACAAACACTCTCTCAGACTCAGTCAAT 4146
Db 213 uLeuSerLysCysLysArgThrGluIleIleThrAsnTyTrpLeuSerAspHisSerAlaI1 233
QY 4147 CTATTAGAACTCAGAAATTAAGAACTCACTCAAAATATCAACACTTACATGAAACTGAA 4206
Db 233 eLysLeuGluLeuArgIleLysAsnLeuThrGlnSerArgSerThrTrpLysLeuAs 253
QY 4207 CAACCTGCTCTCAATGACTACTTGGTGAATAATAACAAAATGAAGGCAAAATAAAGATGT 4266
Db 253 nAsnLeuLeuAsnAspTyTrpValHisAsnGluMetLysAlaGluIleLysMetPh 273
QY 4267 CTTTGAACCAATGAGAACAAAGACACAATGTACCAAGATCTCTGGGGCATATTAAAGC 4326
Db 273 ePheGluThrAsnGluAsnLysAspThrThrTyTrpGlnAsnLeuTrpAspAlaPheLysAl 293
QY 4327 AGTGTGTAGAGGAAATTTATAGCACTAGATGCCTACAGAGAGAAAGACGAAATATCTAA 4386
Db 293 aValCysArgGlyLysPheIleAlaLeuAsnAlaHisLysArgLysGlnGluArgSerLy 313
QY 4387 AATAGACACTTTAATCATCAATTAAGAAGACTAGAGAAAGAGCAACAAATTCAAA 4446
Db 313 sIleAspThrLeuThrSerGlnLeuLysGluLeuGluLysGlnGluGlnThrHisSerLy 333
QY 4447 ACCTAGCAGACAAAGAAATAACTTAAGTATCAGACGAACTGAAGGAGATAGACACA 4506
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

333 Db sAlaSerArgGlnGluIleThrLysIleArgAlaGluLeuLysGluIleGluThrGI 353
4507 QY AAAAGCCCTCAATAATCAATCAATCAAGGAGCTGTTTGTGAAGATCAGCAAAA 4566
353 Db nLysThrLeuGlnLys-IleAsnGluSerArgSerTrpPheGluArgIleAsnLysI 373
4567 QY T-----AGACCACAGACAGACTAATAAAGAAAGAAAGAGAGAAATCAAGAGATCAA 4622
373 Db leAspArgProLeuAlaArgLeuIleLysLysLysArgGluLysAsnGlnIleAspThrI 393
4623 QY TAAAAATGATAAAGGGGATATCACACCCGATCCACAGAAATACAAACTATTATCAGAG 4682
393 Db leLysAsnAspLysGlyAspIleThrAspProThrGluIleGlnThrIleArgG 413
4683 QY AATATTATAAACACCTCTATGCAAAATAAATACTAGAAATCTAGAGAAATGATGATAATCC 4742
413 Db lutyTyrlLysHisLysLeuTyrlAlaAsnLysLeuGluAsnLeuGluGluMetAspThrPheL 433
4743 QY TGGACACATATGTAGCCTGTATGGACCTTGGGGGACAGAACAAAAAGGGGTGAATGCAGA 4802
433 Db euAsp----- 434
4803 QY AATAAAGACAAAGACAAAGAGTAGTGTGGAGTAGGGGTGAGGGGGCAACTTGCCTC 4862
434 Db ----- 434
4863 QY TAATGGACAAGGCCCTGAGCTTTACACCACCCTCTGTATTATTAGGCAAAAGAGATAG 4922
434 Db ----- 434
4923 QY CGAGAGGGTGAGTTGGAGAAGAGGTCAGCTGTAGGTCGAGAGTAGGCCTGCAAGACTG 4982
434 Db ----- 434
4983 QY CATTCTCAACAATAGGCTCTAGATGTCCAGTAGATACCTCAAGGAGCCAGTGCCAG 5042
-434 Db ----- 434
5043 QY GGAGTGATGGCCCTCAGCAAAACCTTCTAGGGCAGGCACAGAGTAAGTTTGCCCACTTC 5102
434 Db ----- 434
5103 QY TGTATTACGATAAACAGTTTGTCTGTGTGATCAAGTAGCCTCCAGTGGAAATGCTGAGTTG 5162
434 Db ----- 434
5163 QY GTCATGTCCCTTTGGCCCTTTTGGCTCCCAAAACACATACACCCCTCTCAAGACTAAAC 5222
435 Db ----- 435
5223 QY AGGAAGAAGTCAATCCCTGAATATACAGCTTAACAGTTCTTAAATTTGAAGCAGTAATTG 5282
443 Db lngluGluValGluSerLeuAsnArgProIleThrGlyAlaGluIleValAlaIleIle 463
5283 QY ATAGCCTTACCAACCAAAAAAGTCCAGGACGAGGATTCACAGCCAAATTTCTACAGA 5342
463 Db snSerLeuSerThrLysLysSerProGlyProAspGlyPheThrAlaGluPheTyrlGlnA 483
5343 QY GGTACAAAGAGAGCTGGTACTATTCTCTTGAACATTTCCAAAAAATAGAA---AATG 5399
483 Db rgyTyrlLysGluGluValProPhePheLeuLysLeuPheGlnSerIleGluLysGlu- 502
5400 QY GGAATCCTCCCTAACTCATTTTACGAGGCCAGCATCATCTGATACCAAAACCTTAGCAGT 5459
503 Db GlyIleLeuProAsnSerPheTyrlGluAlaSerIleIleLeuIleProLysProGlyArg 522
5460 QY GACACACAAAAAGAGAAATTTACGGCCCATATCCCTGATGAACATTTGATGTCAAAATC 5519
523 Db AspThrThrLysLysGluAsnPheArgProIleSerLeuMetAsnIleAspAlaLysIle 542
5520 QY CTCATTAATAATCTGGCAACCAAAATCCACGACGACATCAAAAGCTTATCTACCATGAT 5579
543 Db LeuAsnLysIleLeuAlaAsnGlnIleGlnHisIleLysLysLeuIleHisAsp 562

5580 QY CAAGTTGGCTCATCTCCCTGGCATGCAAGCTGGTTCAAAATATGCAAAATCAATAAATGTA 5639
563 Db GlnValGlyPheIleProGlyMetGlnGlyTrpPheAsnIleArgLysSerIleAsnVal 582
5640 QY GGCCATCACATAACAGAACCAATGACAAAAACACATGATTATCTCAATAGATGCAAGAA 5699
583 Db IleGlnHisIleAsnArgAlaLysAspLysAsnHisMetIleValSerIleAspAlaGlu 602
5700 QY AAGCCCTTTCTCAAAATTTCAACAGCCCTTCTGCTAAATTTCTCAGTAACATAGGTATC 5759
603 Db LysAlaPheAspLysIleGlnProPheMetLeuLysThrIleAsnLysIleGlyIle 622
5760 QY GATGGAAATGATCTCAAAATAAAGAGCTATTATATAC-AAACCCACAGCCAAATATCAT 5818
623 Db AspGlyMetTyrlPheLysIleIleArgAlaIleTyrlAspLysThrThrAlaAsnIleIle 642
5819 QY CTGAATGGGCAAAACCTGGAAGCATTCCTTTGAGAACTGGCACAAGCAAGGATGCCT 5878
643 Db LeuAsnGlyGlnLysLeuGluAlaPheProLeuLysThrGlyThrArgGlnGlyCysPro 662
5879 QY CTCTCACCACCTCTATTCAAGATACTATTGGAAGTTCTGGCAGGCAATCAGGCAATAG 5938
663 Db LeuSerProLeuLeuPheAsnIleValLeuGluPheLeuAlaArgAlaIleArgGlnGlu 682
5939 QY AAAGAAATAAAGGTATTCAAAATAGAAAGAGAGAGTCAATATTGTTCTCTGTTTGCAGAT 5998
683 Db LysGluIleLysGlyIleGlnLeuGlyLysGluGluValLysLeuSerLeuPheAlaAsp 702
5999 QY GACATGTTGTATATTAGAAAACCCCATCTCTCTCAGGCCAAACAACTCTTAACTGATA 6058
703 Db AspMetIleValAspLeuGluAsnProIleValSerAlaGlnAsnLeuLeuLysLeuIle 722
6059 QY AGCAACTTCAGCAAGTCTCAGGACACAAATCAATGTGCAAAATCACAAAGCATTTCTTA 6118
723 Db SerAsnPheSerLys***SerGlyTyrlLysIleAsnValGlnLysSerGlnAlaPheLeu 742
6119 QY TAGCCCAATTAATAGCAAAACAGAGAGCAAAATCATGAGTGAACATCTCATTCACAAATGCT 6178
743 Db TyrlThrAsnAsnArgGlnThrGluSerGluIleMetSerGluLeuProPheThrIleAla 762
6179 QY ACAAGAGATAAATAACCTTAGGAATACAACTTACAAGGACGACGTAGGAACTTCTCAAG 6238
763 Db SerLysArgIleLysTyrlLysGlyIleGlnLeuThrArgAspValLysLeuPheLys 782
6239 QY GAGAACTACAAACCACTGATCAAGGAATAAGAGAGACACAAACAAATGGAACAAAT 6298
783 Db GluAsnTyrlLysProLeuLeuLysGluIleLysGluAspThrLysLysTrpLysAsn*** 802
6299 QY CCATGCTCACAGATAGTAAGAATCAT-----GAAATGCCATACCTGCCCAAGTAAT 6351
803 Db ProCysSerTrpValGlyArgIleAsnIleValLysMetAlaIleLeuProLysValIle 822
6352 QY TATAGATTGAGTGTCTCCCACTCAAGCTACCATGACTTCTTCTTACAGAAATTTGAAAAA 6411
823 Db TyrlArgPheAsnAlaIleProIleLysLeuProMetThrPhePheThrGluLeuGluLys 842
6412 QY ACACATTTAAATTTTATATGGAACCAAAAGAGGCCACAGAGCCCAAGCAATCTTAAG 6471
843 Db ThrThrLeuLysPheIleTrpAsnGln-LysArgAlaArgIleAlaLysSerIleLeuSe 862
6472 QY CAAAAAGACAAAGCTGGAGGTATCATCTACCTGACTTAAACATATACTATAAGGTAC 6531
862 Db rGlnLysAsnLysAlaGlyIleThrLeuProAspPheLysLeuTyrlLysAlaTh 882
6532 QY AGTAACCAAAACCTGATGCTGCTACCAACAAACAGATATATAGACCAATTTGGAACAA 6591
882 Db rValThrLysThrAlaTrpTyrlTyrlGlnAsnArgAspIleAspGlnTrpAsnArgTh 902
6592 QY AGAGACCTCAGAAATTTACACT-GCAATCTACATCCATCTCATCTTTTACAAACCTGACAA 6650
902 Db rGluProSerGluIleThr***LeuThrTyrlAsnTyrlLeuIlePheAspLysProGluTy 922

6651 AAACAAGCAATGGAAGAAAGGATTCCTCTATTATAAATGGTGTGGAAAGAACTGGCTAGC 6710
 Db sasnlyGlnTrpGlyLysAspSerLeuPheAsnLysTrpCysTrpGluAsnTrpLeuAl 942
 6711 CATATGCAGAAAGCTGAAGCTGATCCCTTCCTACACCTTATACAAAAGTTAACTCAAG 6770
 Db alleCysArgLysLeuLysValAspProPheLeuThrProTyrThrLysIleAsnSerAr 962
 6771 ATGAATTAAGACTTAAATATAAGACATAAACCAATAAAACCCCA-GAAGAAAACTTAGG 6829
 Db gtrpIleLysAspLeuAsnValArgProLysThrIleLysThrLeuGluGluAsnLeuGl 982
 6830 CAATACCATTCAGGATATGAGCATGGGCAAGACATTCATGACTAAACACCAAGCAAT 6889
 Db yileThrIleGlnAspIleValGlyLysAspPheMetSerLysThrProLysAlaMe 1002
 6890 GGCACAAAGCCAAATAGACAGTGGGATCTGATTAACCTATAGAGCTTCGACAGC 6949
 Db talathrLysAlaLysIleAspLysTrpAspLeuIleLysLysSerPheCysThrAl 1022
 6950 AAAAAAAACTGCTCATCAGAGTGAACAACCAACCTACAGAATGGGAGAAAAATTTTGC 7009
 Db a-LysGluThrThrIleArgValAsnArgGlnProThrThrTrpGluThrIlePheThr 1042
 7010 TCTATCGATCTGCAAAAGGCTAATATCCAGAGATCTACGAGAAGCTTAAACAAATTTACA 7069
 Db hrTyrSerSerAspLysGlyLeuIleSerArgIleTyrAsnGluLeuLysGlnIleTyrL 1062
 7070 AGAAAAA---AACAAACCCGTCAAATATGGCAAGAGTATGAGCAGACACTTCTCAA 7125
 Db yLysLysThrAsnAsnProIleLysLysTrpPalaLysAspMetAsnArgHisPheSerL 1082
 7126 AAGAAGACATTTATGCAGCAACAACATATCAAAAAACCCATCATTCATTTGGTGTGA 7185
 Db ysGluAspIleThrAlaLysLysHisMetLysLysCysSerSerSerLeuAlaIleA 1102
 7186 GAGAAATGCAAAACCAACACAGTACATACCATCTCATGCTAGTGTAGATGGTGATCA 7245
 Db rgluMetGlnIleLysThrThrMetArgTyrHisLeuThrProValArgMetAlaIleI 1122
 7246 CTAAGAATGCAGAAACACAAATGCTGGAGAGGATGGGAGAAATAGAACACTTTTCC 7305
 Db leLysLysSerGlyAsnAsnArgCysTrpArgGlyCysGlyGluIleGlyThrLeuLeuH 1142
 7306 ACTGTGTGGGAATGTAATAGTTCACCACTTGTGGAAGACAGTGGAGATTCCTTA 7365
 Db isCysTrpTrpAspCysLysLeuValGlnProLeuTrpLysSerValTrpArgPheLeuA 1162
 7366 AGGATCTAGAACCAAGATATCATTTGACCCAGCAATCCCATCTAGTATATATACCAA 7425
 Db rgAspLeuGluLeuGluIleProPheAspProAlaIleProLeuLeuGlyIleTyrProA 1182
 7426 AGGAATATAATCATCTTATATAAGACATGACACATATGTTTATTGACGACACTGA 7485
 Db snAspTrpLysSerCysCysTrpLysAspThrCysThrArgMetPheIleAlaLeuP 1202
 7486 TCACAAATAGCAAGACTTGGAAACCAACCAATGTCATCAGTATAGACTGGATAAAGA 7545
 Db heThrIleAlaLysThrTrpAsnGlnProLysCysProThrMetIleAspTrpIleLysL 1222
 7546 AAACATGCGACATATACACCATGAAATACTATGACGCCATAAAAG-GATGAGTTTCATGT 7604
 Db ysmetTrpHisIleTyrThrMetGluTyrTyrAlaAlaIleLysAsnAspGluPheMetS 1242
 7605 CTTTTCAGAGATATGGATGAGCTGGAACCATCATCTTCAGCAAACTTAACACAAAGAC 7664
 Db erPheValGlyThrTrpMetLysLeuGluThrIleLeuSerLysLeuSerGlnGluG 1262
 7665 AGAAAAACCAACACACATGTTCTCAGTTGTAAAGTGGGAGT 7705
 Db InLysThrLysHisArgIlePheSerLeuIleGlyGlyAsn 1275

RESULT 6

GNLRL1

retrovirus-related reverse transcriptase pseudogene - slow loris
 C:Species: Nycticebus coucang (slow loris)

C>Date: 31-Mar-1988 #sequence_revision 04-Jan-1996 #text_change 14-May-1999

C:Accession: B25313

R:Hattori, M.; Kuhara, S.; Takenaka, O.; Sakaki, Y.

Nature 321, 625-628, 1986

A:Title: L1 family of repetitive DNA sequences in primates may be derived from a sequ

A:Reference number: A93381; MUID:86230917; PMID:2423883

A:Accession: B25313

A>Status: conceptual translation of pseudogene

A:Molecule type: DNA

A:Residues: 1-1260 <HAT>

A>Note: this sequence was constructed from an alignment of six sequences, determined

C:Keywords: reverse transcriptase; pseudogene

Alignment Scores: 6,16e-282 Length: 1260
 Pred. No.: 3658.50 Matches: 743
 Score: 66.90% Conservative: 207
 Percent Similarity: 52.32% Mismatches: 303
 Best Local Similarity: 6.99% Indels: 174
 Query Match: 4 Gaps: 9
 DB:

US-10-083-853b-2 (1-29921) x GNLRL1 (1-1260)

Qy 3434 GGATCAAAATTCACACATATAATATAACCTTAAATGTAATGGCTAAATCCCAATT 3493
 Db 2 GlyLeuSerLysGlyLeuSerIlePheSerIleAsnValAsnGlyLeuAsnCysProLeu 21
 Qy 3494 AAAAGACACAGACTGGCAAAATTTGGATAAGAGTCAAGACCCCATCATGTGCTGTATTCA 3553
 Db 22 LysArgHisArgLeuAlaAspTrpIleGlnLysLeuLysProAspIleCysCysIleGln 41
 Qy 3554 GAGGCCCATCTCACATGATGAAAGACACATAGCTGCTCAAAATAAAGGATGGAGGAGATT 3613
 Db 42 GluSerHisLeuThrLeuLysAspLysTrpArgLeuLysValLysGlyTrpSerSerIle 61
 Qy 3614 TACCAAGTAAATGGTGAACAAAAAAGAGAGGGTTCGAATCTCTAGCTCTCATATAA 3673
 Db 62 PheGlnAlaAsnGlyLysGln---LysLysAlaGlyIleAlaIleLeuPheAlaAspAla 80
 Qy 3674 ACAGACTTTAAACCAACAAAGATCAAAAGACACAAAGAGCCCATACATAATGTTAAG 3733
 Db 81 IleGlyPheLysProThrLysIleArgLysAspLysAspGlyHisPheIlePheValLys 100
 Qy 3734 GCATCAATGGACAAGAGAGCTACTATCTTAATATACATGACCCCAATACAGGACA 3793
 Db 101 GlyAsnThrGlnTyrAspGluIleSerIleIleAsnIleTyrAlaProAsnHisAsnAla 120
 Qy 3794 CCCAGATTCAATAAGCAAGTTCCTTAGAGACCTTACAAAGAGAGCTTTGACTCCCAACAATA 3853
 Db 121 ProGlnPheIleArgGluThrLeuThrAspMetSerAsnLeuIleSerSerThrSerIle 140
 Qy 3854 ATAGTGGAGTCTAAATAATAAATAGACACTTTTAACACCCCACTGCCAATATTAGGACA 3913
 Db 141 ValValGly-----AspPheAsnThrProLeuAlaValLeuAspArg 154
 Qy 3914 TCAATGACACAGAAAAATAACAAGGATATCCAGGAGTTGAACCTGAGCTCTGGACCAACCG 3973
 Db 155 SerSerLysLysLysLeuSerLysGluIleLeuAspLeuAsnSerThrIleGlnHisLeu 174
 Qy 3974 GACCTAATAGATATCTACAGAACTCCCAACCCCAATCAACAGAAATATACACTCTCTCA 4033
 Db 175 AspLeuThrAspIleTyrArgThrPheHisProAsnLysThrGluTyrThrPhePheSer 194
 Qy 4034 GCATCACAATTACACCTATTTTAAATTTGACCATGTAATTTTAAGTAAACACTCTCTCAGC 4093
 Db 195 SerAlaHisGlyThrTyrSerLysIleAspHisIleLeuGlyHisLysSerAsnLeuSer 214
 Qy 4094 AAATGCCAAAGAACAGAAATCTTAACAAACAGCTCTCTCAGACTACAGTGCATCTATTATA 4153
 Db 215 LysPheLysLysIleGluIleIleProCysIlePheSerAspHisHisGlyIleLysVal 234


```
Db      804  erTrpLeuGlyArgIleAsnIleValLysMetSerIleLeuProLysAlaIleTyrAsnP 824
QY      6359 TCAGTGTACCCCAAGCTTACCATTGACCTTCTTCCACAGAAATGGAAAAACAACATT 6418
Db      824  heAsnAlaIleProLysAlaProLeuSerIlePheLysAspLeuGluLysIleIleL 844
QY      6419 TAAATTTTCATATGGAAACCAAAAAAGAGCCACAGAGCCCAAGCAACATCTTAAGCAAAAAG 6478
Db      844  euHisPheIleIleTrpAsnGlnLysLys-ProGlnIleAlaLysThrLeuLeuSerAsnLys 863
QY      6479 ACAAAAGCTGGAGGTATCATGCTACTGCTTAAACTATACTATAAGCTTACAGTAACC 6538
Db      864  AsnLysAlaGlyIleThrLeuProAspLeuArgLeuTyrTyrLysSerIleValIle 883
QY      6539 AAAAAGCTGATGCTACTGTTACCAAAACAGATATATAGACCAATGGACAGACAGACACC 6598
Db      884  LysThrAlaIleTrpIleHisLysAsnArgGluValAspValTrpAsnArgIleGluAsn 903
QY      6599 TCAGAAATATAC-ACAGCAATCTACATCCATCTGATCTTGTACAAAACCTGACAAAACAAG 6657
Db      904  GlnGluMetAspProAlaThrTyrHisTyrLeuIlePheAspLysProIleLysAsnIle 923
QY      6658 CAATGGAAAAGGATTCCTATTATTAATAATGGTGTGGAAAACCTGGCTAGCCATATGCC 6717
Db      924  GlnTrpGlyLysAspSerLeuPheAsnLysTrpCysTrpValAsnTrpLeuAlaIleCys 943
QY      6718 AGAAAGCTGAAACTGGATCCCTCTTACACCTTATACAAAGTTAACTCAAGATGAATT 6777
Db      944  ArgArgLeuLysLeuAspProHisLeuSerProLeuThrLysIleAspSerHisTrpIle 963
QY      6778 AAAGACTTAAATATAGACATAAAACCAATAAAACCCCA-GAAGAAAACCTAGGCAATACC 6836
Db      964  LysAspLeuAsnLeuArgHisGlnThrIleLysIleLeuGluLysSerAlaGlyLysThr 983
QY      6837 ATTCAGGATATGGACATGGGCAAGACCTTCATGACTAAACCAACCAAGCAATGCCACA 6896
Db      984  LeuGluGlyIleSerLeuGlyLysTrpPheMetArgTrpGlnAlaIleGluAla 1003
QY      6897 AAAGCAAAATAGACAAGTGGGATCTGATTAACATATAGAGCTTCTGCACAGCAAAAAA 6956
Db      1004 ValSerLysIleHisTyrTrpAspLeuLysLysLeuLysSerPheCysThrAla-LysAs 1023
QY      6957 AACTGTATCATGAGTGAACAAGCAACCTACAGAAATGGGAGAAATTTTTCGAATCATCG 7016
Db      1023 nileValSerLysAlaSerArgGlnProSerGlnProSerGlnLysIlePheAlaGlyTy-Th 1043
QY      7017 ATCTGACAAGCGTAAATATCCAGACATCTACGAACACTTAACAAATTTACAGAAAAA 7076
Db      1043 rSerAspLysGlyLeuIleThrArgIleHisArgGluLeuLysHisIleAsnLysLysAr 1063
QY      7077 AACAAAC---CCCGTCAAAATATGGCAAGAGATATAGCAGACACTTCTCAAAAAGA 7132
Db      1063 gThrArgAspProIleSerGlyTrpAlaArgAspLeuLysArgAsnPheSerLysGluAs 1083
QY      7133 CATTTATCAGCAACAACATATGAAAAAACCTCATCATCTGCTGCTGTAGAGAAT 7192
Db      1083 pArgHisThrIleTyrLysHisMetLysLysSerSerSerLeuIleIleArgGluLe 1103
QY      7193 GCAAAACAACACCATGACATACCATCTCATGCTAGTGTAGATGGTGATCACTAAAAA 7252
Db      1103 tGlnIleLysThrThrLeuArgTyrHisLeuThrProValArgValAlaHisIleThrLy 1123
QY      7253 GTCAGGAACAACAATGCTGGAGAGGATGTGGAGAAATAGAACACTTTTCCACTGTGTG 7312
Db      1123 sSerProAsnGlnArgCysTrpArgGlyCysGlyLysGlyThrLeuLeuHisCysTr 1143
QY      7313 GTGGGAATGTAATATAGTTCACCAATCTGGAAGACAGCTGTGAGATTCCTTAAGGATCT 7372
Db      1143 pTrpGluCysProLeuIleArgSerPheTrpLysAspValTrpArgIleLeuArgAspLe 1163
QY      7373 AGAACCAAAATATCATTTGACCCAGCAATCCCATTTAGTATATATACCAAGGAATA 7432
      1:::  ::::  |||||||  |||||||  |||||||  |||||||  |||  |
```

```
Db      1163 uLysIleAspLeuPropheAspProIleIleProLeuLeuGlyLeuTyrProGluAspGl 1183
QY      7433 TAAATCATCTTATTATAAGACACATGCACATATGTTATTATGCAGCACTGATCACAAT 7492
Db      1183 nLysSerGlnTyrAsnLysAspIleCysThrArgMetPheIleAlaIleGlnPheIleI 1203
QY      7493 AGCAAAAGACTTGGAAACCAACCCAAATGTCCATCAGTATGATGATGAATAAGAACATG 7552
Db      1203 eAlaLysSerTrpLysLysProLysCysProSerThrHisGluTrpThrSerLysLeuTr 1223
QY      7553 GCACATATACACCATCAATACTATGACGCCATAAAAGGAT- ---GAGTTTCATGTCCTT 7608
Db      1223 pTyrMetTyrThrMetGluTyrTyrAlaAlaLeuLysLysAspGlyAspPheThrSerPh 1243
QY      7609 TCACAGATATGATGAGTGAAGCTGGAACCAACCATCTCTCAGCAAACTAAACAA 7660
Db      1243 eMetPheThrTrpMetGluLeuGluHisIleLeuLeuSerLysValSerGln 1260

RESULT 7
GNMSLL
retrovirus-related reverse transcriptase homolog - mouse retrotransposon
N:Alternate names: L1md repetitive element ORF-2; LINE-1 hypothetical protein; ORF 39
C:Species: Mus musculus (house mouse)
C:Date: 31-Mar-1989 #sequence_revision 08-Jan-1999 #text_change 16-Jun-2000
C:Accession: B58927; B24906; I49130; A23430
R:Loeb, D.D.; Padgett, R.W.; Hardies, S.C.; Shehee, W.R.; Comer, M.B.; Edgell, M.H.;
Mol. Cell. Biol. 6, 168-182, 1986
A:Title: The sequence of a large L1md element reveals a tandemly repeated 5' end and
A:Reference number: A39072; MUID: 87064284; PMID: 3023821
A:Accession: B58927
A:Status: translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-1281 <LOE1>
A:Cross-references: GB:M13002; NID:g200849; PIDN:AAA66024.1; PID:g804788
A:Note: sequence constructed using the first potential start codon for ORF2
A:Accession: B24906
A:Molecule type: DNA
A:Residues: 'NNQESHSTNOKEDSHKNR', 1-1281 <LOE2>
A:Cross-references: GB:M13002; NID:g200849
A:Note: sequence shown in Fig. 2
R:Martin, S.L.; Martin, S.L.
Gene 153, 261-266, 1995
A:Title: Characterization of a LINE-1 cDNA that originated from RNA present in ribonu
A:Reference number: I49129; MUID: 95180729; PMID: 7533116
A:Accession: I49130
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-85, 'L', 87-358, 'K', 360-706, 'R', 708-735, 'A', 737-760, 'W', 762-927, 'D', 929-1
A:Cross-references: EMBL:U15647; NID:g558906; PIDN:AAA67727.1; PID:g558908
R:Mottez, E.; Rogan, P.K.; Manueldis, L.
Nucleic Acids Res. 14, 3119-3136, 1986
A:Title: Conservation in the 5' region of the long interspersed mouse L1 repeat: impl
A:Reference number: A23772; MUID: 86176789; PMID: 3008107
A:Accession: A23772
A:Molecule type: DNA
A:Residues: 'NNQESHSTNOKEDSHKNR', 1-245, 'K', 247-423, 'SYTQQNWKTKWTN', 439, 'WTDTRYOS'
A:Cross-references: GB:X03725; NID:g52829; PIDN:CAA27363.1; PID:g1334115
C:Superfamily: pol polypeptide
C:Keywords: reverse transcriptase

Alignment Scores:
Pred. No.: 1.67e-273 Length: 1281
Score: 3552.50 Matches: 729
Percent Similarity: 65.39% Conservativeness: 208
Best Local Similarity: 50.87% Mismatches: 329
Query Match: 6.79% Indels: 174
DB: 1 Gaps: 9

US-10-083-853B-2 (1-29921) x GNMSLL (1-1281)
QY      3434 GGATCAATTCACACATAATAATATTAACCTTAATGCTAAATGGCTAAATTCCTCCCAATT 3493
Db      10  GlySerAsnAsnTyrPheSerLeuIleSerLeuAsnIleAsnGlyLeuAsnSerProIle 29
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
```

Qy	3494	AAAAGACACAGACTGGCAAAATTGGATAAAGAGTCAAGACCCATCAGTGTCTGTATTTCAG	3553
Db	30	LysArgHisArgLeuThrAspTrpLeuHisLysGlnAspProThrPheCysCysLeuGln	49
Qy	3554	GAGCCCATCTCACATCAAAAGACACACATAGCTCAAAATAAAGGATGAGGAGAGATT	3613
Db	50	GlnThrHisLeuArgGluLysAspArgHisIleuArgValLysGlyTrpLysThrIle	69
Qy	3614	TACCAACTAAATGGAACCAAAAAAAGACGAGGGGTGCAATCCCTAGTCTCTGATAAA	3673
Db	70	PheGlnAlaasnGly---LeuLysLysGlnAlaGlyValAlaIleLeuIleSerAspLys	88
Qy	3674	ACAGCTTTAAACCAACAAGATCAAAAGAGACAAAGAGCCCATTAATAATGGTAAAG	3733
Db	89	IleAspPheGlnProLysValIleLysLysAspLysGluGlyHisPheIleLeuIleLys	108
Qy	3734	GCATCAATGGNACAAAGAGCTAACTATCTCTAAATATATATACGCCCAATACAGGACA	3793
Db	109	GlyLysIleuGlnGluLeuSerIleLeuAsnIleTyAlaProAsnAlaArgAla	128
Qy	3794	CCCAGATTCAATAAGCAAGTCTCTAGAGACCTACAAAGAGACTTTGACTCCACACAATA	3853
Db	129	AlaThrPheIleArgAspThrLeuValLysLeuLysAlaTyIleAlaProHisThrIle	148
Qy	3854	ATAGTGGAGTCTAAATAATAATAGACACTTTAAACACCCACTGCCCCAATATTAGGCAGA	3913
Db	149	IleValGly-----AspPheAsnThrProLeuSerSerLysAspArg	162
Qy	3914	TCATGAGACAGAAAATTAACAAGGATATCCAGGAGTTGAAGTCTGGACCAACGC	3973
Db	163	SerTrpLysGlnLysLeuAsnArgAspThrValLysLeuThrGluValMetLysGlnMet	182
Qy	3974	GACCTAATAGATATCTACAGAACTCCACCACCCCAAAATCAACAGAAATATACACTCTCTCA	4033
Db	183	AspLeuThrAspIleTyArgThrPheTyProLysThrLysGlyTyThrPhePheSer	202
Qy	4034	GCATCACATACACTATTTTAAATAGACCATGTAATTTTAAAGTAAACACTCTCTCAGC	4093
Db	203	AlaProHisGlyThrPheSerLysIleAspHisIleGlyHisLysThrGlyLeuAsn	222
Qy	4094	AAATGCAAAAGACAGAAATCCTTAACAACAGCTCTCAGACTACAGTCAATCTATTTA	4153
Db	223	ArgTyrlLysAsnIleGluIleValProCysIleLeuSerAspHisHisGlyLeuArgLeu	242
Qy	4154	GAACCTCAGAATTAAAGAACTCACTCAAAATCACAACTACATGGAACCTGAACAACTG	4213
Db	243	IlePheAsnAsnAlaAsnAsnGlyLysProThrPheThrTrpLysLeuAsnAsnThr	262
Qy	4214	CTCCTGAATGACTACTGGGTAAATAAACAATAAGAGCAAAATAAAGATGTTCTTTGAA	4273
Db	263	LeuLeuAsnAspThrLeuValLysGluGlyIleLysLysGluIleLysAspPheLeuGlu	282
Qy	4274	ACCAATGAGAAAGACACATGTCACAGATCTCAGAACTCTGGGGCATATTTAAAGCAGTGT	4333
Db	283	PheAsnGluAsnGluAlaThrThrTyProAsnLeuTrpAspThrMetLysAlaPheLeu	302
Qy	4334	AGGGGAAATTTATAGCAGTGTGCTTACAGAGAAAGAGGAGAAATATCTAAATAGAC	4393
Db	303	ArgGlyLysLeuIleAlaLeuSerAlaSerLysLysLysArgGluThrAlaHisThrSer	322
Qy	4394	ACCTTAACATCACAATTTAAAGACTAGAGAAAGAGCAAAACAAATTTCAAAGCTGAGC	4453
Db	323	SerLeuThrThrHisLeuLysAlaLeuGluLysLysGluAlaAsnSerProLysArgSer	342
Qy	4454	AGAAGACAGAAATACTAGATCAGACAGCAACTGAAGGAGATAGACACAAAAAGCC	4513
Db	343	ArgArgGlnGluIleLysLeuArgGlyGluIleAsnGlnValGluThrArgThr	362
Qy	4514	CTTCAATAATCAATCAATCCAGGCTGGTTTTTTTGAAGAGATCAGCAAAATAGA---	4570
Db	363	IleGlnArg-IleAsnGlnThrArgSerTrpPhePheGluLysIleAsnLysIleAspLys	382

Nucleic Acids Res. 18, 4099-4104, 1990
 A:Title: Selective cloning and sequence analysis of the human L1 (LINE-1) sequences which
 A:Reference number: S23649; MUID:90332398; PMID:216587

A:Accession: S23650
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA

A:Residues: 1-712 <HOH>

A:Cross-references: EMBL:X52235

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1990

C:Genetics:

A:Mobile element: LINE-1

A:Start codon: GTG

C:Superfamily: pol polypeptide

Alignment Scores:
 Pred. No.: 3,85e-220 Length: 712
 Score: 2880.50 Matches: 573
 Percent Similarity: 87.11% Conservative: 49
 Best Local Similarity: 80.25% Mismatches: 87
 Query Match: 5.50% Indels: 11
 DB: 2 Gaps: 2

US-10-083-853B-2 (1-29921) x S23650 (1-712)

Qy 5583 GTTGGCTGCTACCTGGGATGCAAGGCTGTTCAAAATATGCAAAATCAATAATGTAGGC 5642
 Db 1 MetGlyPheIleProGlyMetGlnAspTrpPheAsnMetHisLysSerIleAsnValIle 20
 Qy 5643 CATCACATAACAGAACCAATGACAAAAACACATGATTATCTCAATAGATGCGAAGAA 5702
 Db 21 GlnHisIleAsnArgTrpLysAspLysAsnHisMetIleValSerIleAspAlaGluLys 40
 Qy 5703 GCCTTTCTCAAAATTAACAGCCCTTCATGCTCAAAATTTCTCAGTAACATAGGTATCGAT 5762
 Db 41 AlaPheAspLysIleGlnGlnProPheMetLeuLysThrLeuAsnLysLeuGlyIleAsp 60
 Qy 5763 GGAATGATATCTCAAAATTAAGAGCTATTATATAC - AAACCCACAGCGCAATATCATCTG 5821
 Db 61 GlyThrTrpPheLysIleIleArgAlaIleIleAspLysProThrAlaAsnIleIleLeu 80
 Qy 5822 AATGGGCAAAACCTGGAGCATTCCTTTCAGAACTGGCAGCAAGCAAGGATGCCCTCTC 5881
 Db 81 AsnGlyGlnLysLeuGluAlaPheProLeuLysThrGlyThrArgGlnGlyCysProLeu 100
 Qy 5882 TCACCATCTCTATTCAAGATACTATTGGAAAGTTCTGGCCAGGCAATCAGGCAATAGAA 5941
 Db 101 SerProLeuLeuPheAsnIleValLeuGluValLeuAlaArgAlaIleArgGlnGluLys 120
 Qy 5942 GAATAAAGGATATTCAATAGAAAGAGAGAGTATATGTTCTCTGTTTGCAGATGAC 6001
 Db 121 GluIleLysGlyIleGlnLeuGlyLysGluGluValLysLeuSerLeuPheAlaAspAsp 140
 Qy 6002 ATGTTTCTATATTAGAAACCCCATCTGCTCAGGCCCAAAACCTCTTAAGCTCATAGC 6061
 Db 141 MetIleValTrpLeuGluAsnProIleValSerAlaGlnAsnLeuLysProIleSer 160
 Qy 6062 AACTTCAGCAAGTCTCAGGACACAAAATCAATGTGCAAAATACACAGCATCTCTTATAC 6121
 Db 161 AsnPheSerLysValSerGlyTrpLysIleAsnValGlnLysSerGlnAlaPheLeuTrp 180
 Qy 6122 GCCAATATAGCAAAACAGAGCGCAAAATCATGATGATGATCTCATCTCATTCACAAATGCTACA 6181
 Db 181 ThrAsnAsnArgGlnThrGluSerGlnIleMetAsnGluLeuProPheThrIleAlaSer 200
 Qy 6182 AAGAGATAAATACCTAGGATACAACTTACAGGACACGCTAGGAACTCTTCAAGGAG 6241
 Db 201 LysArgIleLysTrpLeuGlyIleGlnIleThrArgAspValLysAspLeuPheLysGlu 220
 Qy 6242 AACTACAAACACCTGATCAAGGAAATTAAGAGGACACAAACAAATGGAAACATTTCCA 6301
 Db 221 AsnTrpLysProLeuLeuLysGluIleLysGluAspThrAsnLysTrpLysAsnIlePro 240
 Qy 6302 TGCTCACAGATA-----GTAAGAATCATGAAATGCC-ATACTGCCCAAGTAATATAT 6354

Db 241 CysSerTrpValGlyArgIleAsnIleMetLysMetAlaIleLeuProLysValIleTrp 260
 Qy 6355 AGATTCAAGTCTACCCCATCAAGTACCATCTTCTTCACAGATTCGGAAGAAACA 6414
 Db 261 ArgPheAsnThrIleProIleLysLeuProMetThrPhePheTrpGluLeuGluLysThr 280
 Qy 6415 ACTTTAAATTTTATATGGAACCAAAAGAGCCACAGAGCCCAAGCAATCTTAAAGCAA 6474
 Db 281 ThrLeuLysPheIleTrpAsnGln-LysArgAlaArgIleAlaLysAlaIleArgSerGln 300
 Qy 6475 AAAGAACAAAGCTGGAGGTATCATCTACCTGACTTAAACTATATACATATAGGTCTACAGT 6534
 Db 300 nLysAsnLysSerGlyGlyIleThrLeuProAspPheLysLeuTrpLysAlaThrVa 320
 Qy 6535 AACCAAAACCTGCTGGTACTGGTACCAAAACAGATATATACCAAAATGGAACAGAACAGA 6594
 Db 320 lThrLysThrAlaTrpTrpTrpTrpGlnAsnArgAspIleAspGlnTrpAsnArgThrGln 340
 Qy 6595 GACCTCAGAAATTTACACT-GCAATCTACATCTCATCTGATCTTTGACAAACCTGACAAAA 6653
 Db 340 uProSerGluIleThrProHisIleTrpAsnTrpLeuIlePheAspLysProGluLysAs 360
 Qy 6654 CAAGCAATGGAAGGATTCCTTATTTAATAATGTTGTTGGAAAACTGGCTAGCCAT 6713
 Db 360 nGluGlnTrpGlyLysAspSerLeuPheAsnLysTrpCysTrpGluAsnTrpLeuAlaIle 380
 Qy 6714 ATGCAGAAAGCTGAACCTGGATCCCTTCTTACACCTTATATACAAAGTTAACTCAAGATG 6773
 Db 380 eCysArgLysLeuLysLeuAspProPheLeuThrProTrpTrpLysIleAsnSerArgTr 400
 Qy 6774 AATTAAGACTTAATATAGACATAAAACCAATAAAACCCCA-GAAGAAAACTTAGGCAA 6832
 Db 400 pIleLysAspLeuIleValArgProLysThrIleLysThrLeuGluGluAsnLeuGlyIle 420
 Qy 6833 TACCATTGAGATATGACATGGCCAAAGACTTCATGACTTAAACACCAAAAGCAATGGC 6892
 Db 420 eThrIleGlnAspIleGlyMetGlyLysAspPheMetSerLysThrProLysAlaMetAl 440
 Qy 6893 AACAAAAACCAAAATAGACAGTGGATCTGATTAAACTATAGACTTCTCAGCAGCAAA 6952
 Db 440 aThrLysAlaLysIleAspLysTrpAspLeuIleLysLeuLysSerPheCysThrAla-L 460
 Qy 6953 AAAAATCTGTCATCAGAGTGAACCAACCTACAGATGGGAGAAAATTTTTCGAATCT 7012
 Db 460 ySGLuThrThrIleArgValAsnArgGlnProThrLysTrpGluLysThrPheAlaThr 480
 Qy 7013 ATCGATCTGCAAAAGGCTAATATCCAGAGATCTACGAGAACTTAACAAATTTTACAAGA 7072
 Db 480 yrSerSerAspLysGlyLeuIleSerArgIleTrpAsnGluLysGlnIleTrpLysL 500
 Qy 7073 AAAA---AACAAACCCGTCAAAATATGGGCAAGGATATGAGCAGACACTTCTCAAAAG 7128
 Db 500 ysLysThrAsnAsnProIleLysLysTrpAlaLysAspMetAsnArgHisPheSerLysG 520
 Qy 7129 AAGCATTTATGACGCCCAACAAACATATGAAAAAACCTCATCATCTGCTGTAGAG 7188
 Db 520 luAspIleTrpAlaAlaLysLysHisMetLysLysCysSerProSerLeuAlaIleArgG 540
 Qy 7189 AATGCAAAACAAACCCAGCTGACATACCATCTCATCTGCTAGTGTAGATGTGTACTACTA 7248
 Db 540 luMetGlnIleLysThrThrMetArgTrpHisLeuThrProValArgMetAlaIleLe 560
 Qy 7249 AAAAGTCAGGAAACCAAAATGCTGGAGAGATGTTGGGAATAGGACACTTTTCCACT 7308
 Db 560 ysLysSerGlyAsnAsnArgCysTrpArgGlyCysGlyIleGlyIleGlyThrLeuLeuHsc 580
 Qy 7309 GTTGGTGGGATGTAATTTAGTCAACATTGTGGAAGACAGTGTGGAGATTCCTTAAG 7368
 Db 580 ySTrpTrpAspCysLysLeuValGlnProLeuTrpLysSerValTrpArgPheLeuArg 600
 Qy 7369 ATCTAGAACCAAGAAATATCTTTGACCCAGCAATCCCATCTACTGATGATATACCAAGG 7428

R:Dombroski, B.A.
submitted to the EMBL Data Library, January 1992
A:Description: Isolation of an active human transposable element.
A:Reference number: S65823

A:Accession: S65824

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1275 <DDB>

A:Cross-references: EMBL:M80340; NID:g339767; PIDN:AAA51622.1; PID:g339771

C:Superfamily: pol polyprotein

Alignment Scores:
Pred. No.: 3,01e-168 Length: 1275
Score: 2229.00 Matches: 409
Percent Similarity: 95.24% Conservative: 11
Best Local Similarity: 92.74% Mismatches: 21
Query Match: 4.25% Indels: 0
DB: 2 Gaps: 0

US-10-083-853B-2 (1-29921) x S65824 (1-1275)

```
QY 29919 ACTTTCTTACAGAAATGGAATAAAGTCTTAAAGTTCATGTGGACCAAAAAGAGCC 29860
Dbb 835 ThrPhePheThrGluLeuGluLysThrThrLeuLysPheIleTrpAsnGlnLysArgAla 854
QY 29859 CACATCGCCCAAGTCAATCCTAAGCCAAAAGAAAGAAAGCTGGAGGCATCATGTACCTGAC 29800
Dbb 855 ArgIleAlaLysSerIleLeuSerGlnLysAsnLysAlaGlyIleThrLeuProAsp 874
QY 29799 TTCAACTACACTACAGCTCAGTAACCAAAACAAACATGTTAGTGGTACCAAAACAGA 29740
Dbb 875 PheLysLeuTyrlsAlaThrValThrLysThrAlaTrpTyrlsTrpGlnAsnArg 894
QY 29739 GATATCAACCAATGGAACAGACAGAGCCCTCAGAAATATCCCATATCTACACTAT 29680
Dbb 895 AspIleAspGlnTrpAsnArgThrGluProSerGluIleMetProHisIleTyrlsAsnTyrls 914
QY 29679 CTGATCTTTGACAAACCTGAGAAACAAAGCAATGGGAAAGATTCCTATTTAATAAA 29620
Dbb 915 LeuIlePheAspLysProGluLysAsnLysGlnTrpGlyLysAspSerLeuPheAsnLys 934
QY 29619 TGGTCTGAGAAACTGGCTAGCCCTATGTAAGAGCTGAAAGTGAATCCCTTCCTTACA 29560
Dbb 935 TrpCysTrpGluAsnTrpLeuAlaIleCysArgLysLeuLysLeuAspProPheLeuThr 954
QY 29559 CCTTATACAAATTAATCAAGATGGATTAAGACTTAAATCTTAGACTTAAACCTA 29500
Dbb 955 ProTyrlsLysIleAsnSerArgTrpIleLysAspLeuAsnValLysProLysThrIle 974
QY 29499 AAATCCCTAGAAAGAAACCCAGGCAATACCATTCAGGACATAGGATGGCAAGGACTTC 29440
Dbb 975 LysThrLeuGluGluAsnLeuGlyIleThrIleGlnAspIleGlyValGlyLysAspPhe 994
QY 29439 ATGCTTAAACACCAAGCAATGGCAACAAAGCCAAATTCACAAATGGGATCTAAT 29380
Dbb 995 MetSerLysThrProLysAlaMetAlaThrLysAspLysIleAspLysTrpAspLeuIle 1014
QY 29379 AAATAAGAGCTCTGCACAGCAAAAGAAACTACCATCAGAGTGAACAGGCAACTTACA 29320
Dbb 1015 LysLeuLysSerPheCysThrAlaLysGluThrIleArgValAsnArgGlnProThr 1034
QY 29319 GAATGGGAGAAATTTTTCACACCTACTCATTTGACAAAGGGCTAATATCCAGATCTAC 29260
Dbb 1035 ThrTrpGluLysIlePheAlaThrTyrlsSerAspLysGlyLeuIleSerArgIleTyrls 1054
QY 29259 AATGAACCAACAAATTTACAGAAACAAAGAAACCAAGACCCCATCAAAAGTGGTGAAG 29200
Dbb 1055 AsnGluLeuLysGlnIleTyrlsLysLysThrAsnAsnProIleLysLysTrpAlaLys 1074
QY 29199 GATATGACAGACACTTCTCAAGAGAGCATTTATGACGCCCAAAAACACATGAAAGAA 29140
Dbb 1075 AspMetAsnArgHisPheSerLysGluAspIleTyrlsAlaAlaLysLysHisMetLysLys 1094
```

```
QY 29139 TACTCATCATCACTGGCCCATCAGAGAAATGCAAAATCAAAACCAACCAATGAGATACCATCTC 29080
Dbb 1095 CysSerSerSerLeuAlaIleArgGluMetGlnIleLysThrThrMetArgTyrlsLeu 1114
QY 29079 ACACCAAGTTAGATGGCAATCATTTAAAGTTCAGGAAACAAAGAGTGGTGGAGAGAAATGT 29020
Dbb 1115 ThrProValArgMetAlaIleLysLysSerGlyAsnAsnArgCysTrpArgGlyCys 1134
QY 29019 GGAGAAATAGGACACTTTTACACTGTTGGTGGGACTGTCAACTAGTCTCAACCATTTGGTGG 28960
Dbb 1135 GlyGluIleGlyThrLeuLeuHisCysTrpTrpAspCysLysLeuValGlnProLeuTrp 1154
QY 28959 AAGTCAGTGTGGTATTCTCAGGATCTAGAACTAGAAATACCAATACCAATTCACCCAGCCATC 28900
Dbb 1155 LysSerValTrpArgPheLeuArgAspLeuGluLeuGluIleProPheAspProAlaIle 1174
QY 28899 CCATTACTGGGTATATACCCAAAGGATTATAATCATGTGCTATTAAGACACACACACACA 28840
Dbb 1175 ProLeuLeuGlyIleTyrlsProAsnGluTyrlsSerCysCysTyrlsAspThrCysThr 1194
QY 28839 TGTATGTTTATAGCAGCAGTATTCACATAGCAAGAACTTGGAAACCACTTAATCTCCA 28780
Dbb 1195 ArgMetPheIleAlaAlaLeuPheThrIleAlaLysThrTrpAsnGlnProLysCysPro 1214
QY 28779 ACAACGATAGACTGGATTAAAGAAATGTGGCAGATATACACCATGGAATCTATGAGCC 28720
Dbb 1215 ThrMetIleAspTrpIleLysLysMetTrpHisIleTyrlsThrMetGluTyrlsAlaAla 1234
QY 28719 ATAAATATGATGAGTTCCTGCTGCTTTAGGACATGAGTGAAGCTGGAACCATCATTT 28660
Dbb 1235 IleLysAsnAspGluPheIleSerPheValGlyThrTrpMetLysLeuGluThrIleIle 1254
QY 28659 CTCAGCAAACTATCACAAGGACAAACCAACACCGCATGTTCTCACTCATAGTGGG 28600
Dbb 1255 LeuSerLysLeuSerGlnGluGlnLysThrLysHisArgIlePheSerLeuIleGlyGly 1274
QY 28599 AAT 28597
Dbb 1275 Asn 1275
RESULT 11
I38588
Reverse transcriptase homolog - human retrotransposon L1
N:Alternate names: ORF2 protein
C:Species: Homo sapiens (man)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 26-Aug-1999
C:Accession: I38588
R:Holmes, S.E.; Dombroski, B.A.; Krebs, C.M.; Boehm, C.D.; Kazazian, H.H.
A:Title: A new retrotransposable human L1 element from the LRE2 locus on chromosome 1
Nature Genet. 7, 143-148, 1994
A:Reference number: I38587; MUID:95004577; PMID:7920631
A:Accession: I38588
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1275 <RES>
A:Cross-references: EMBL:U09116; NID:g483914; PIDN:AA60345.1; PID:g483916
C:Superfamily: pol polyprotein
Alignment Scores:
Pred. No.: 3,91e-167 Length: 1275
Score: 2215.00 Matches: 408
Percent Similarity: 95.01% Conservative: 11
Best Local Similarity: 92.52% Mismatches: 22
Query Match: 4.23% Indels: 0
DB: 2 Gaps: 0
US-10-083-853B-2 (1-29921) x I38588 (1-1275)
QY 29919 ACTTTCTTACAGAAATGGAATAAAGTCTTAAAGTTCATGTGGACCAAAAAGAGCC 29860
Dbb 835 ThrPhePheThrGluLeuGluLysThrThrLeuLysPheIleTrpAsnGlnLysArgAla 854
QY 29859 CACATCGCCCAAGTCAATCCTAAGCCAAAAGAAAGAAAGCTGGAGGCATCATGTACCTGAC 29800
```

Db	855	ArgIleAlaLysSerIleLeuSerGlnLysAsnLysAlaGlyIleThrLeuProTyr	874
QY	29799	TTCAAACTACACTCAAGGCTACAGTAACCAAAACAACATGGTACTGGTACCAAAAACAGA	29740
Db	875	PhelYsLeuYrTyrLysAlaThrValThrLysThrAlaTrpTyrTrpTyrGlnAsnArg	894
QY	29739	GATATCAACCAATGAACAGACAGCCCTCAGAAATAATAGCCGATATCTACAACATAT	29680
Db	895	AspIleAspGlnTrpAsnArgThrGluProSerGluIleMetProHisIleTyrAsnTyr	914
QY	29679	CTGATCTTTGACAAACCTGAGAAACCAACCAATGGGAAAGGATCCCTATTATAATAA	29620
Db	915	LeuIlePheAspLysProGluLysAsnLysGlnTrpGlyLysAspSerLeuPheAsnLys	934
QY	29619	TGGTGCTGAGAAACTGGCTAGCCCTATGTAGAAAGCTGAAACTGGATCCCTCTTACA	29560
Db	935	TrpCysTrpGluAsnTrpLeuAlaIleCysArgLysLeuAspLeuPheLeuThr	954
QY	29559	CCTTATACAAAATAATTCAGATGATTAAAGACTTAAATGTAGACTTAAACCATTA	29500
Db	955	ProTyrThrLysIleAsnSerArgTrpIleLysAspLeuAsnValLysProLysThrIle	974
QY	29499	AAATCCCTAGAAACCCAGGCAATACCATTCAGACATAGCGATGGGCAAGGACTTC	29440
Db	975	LysThrLeuGluGluAsnLeuGlyIleThrIleGlnAspIleGlyValGlyLysAspPhe	994
QY	29439	ATGCTCTAAACACCAACCAAGCAATGGCAACAAAGCCAAATAATGCACAAATGGGATCTAAT	29380
Db	995	MetSerLysThrProLysAlaMetAlaThrLysAspLysIleAspLysTrpAspLeuIle	1014
QY	29379	AAACTAAGAGCTTCTGCACAGCAAAAGAACTACCATCAGAGTGAACGCAACTTACA	29320
Db	1015	LysLeuLysSerPheCysThrAlaLysGluThrIleArgValAsnArgGlnProThr	1034
QY	29319	GAATGGGAGAAATTTTGGCACTACTCATTTGACAAAGGCGTAAATATCCAGAACTAC	29260
Db	1035	ThrTrpGluLysIlePheAlaThrTyrSerAspLysGlyLeuIleSerArgIleTyr	1054
QY	29259	AATGAACCTCAACAAATTTACAGAAACCAAGCAAGCCCATCAAAAGTGGGTGAAG	29200
Db	1055	AsnGluLeuLysGlnIleTyrLysLysThrAsnAsnProIleLysLysTrpAlaLys	1074
QY	29199	GATATGAACAGACACTTCTCAAAAGAGGCAATTTATGCAGCCCAAAACACATGAAAAA	29140
Db	1075	AspMetAsnArgHisPheSerLysGluAspIleTyrAlaAlaLysHisMetLysLys	1094
QY	29139	TACTCATCATCTGGCCATCAGAGAAATGCAATCAAAACCAACATGATACCATCTC	29080
Db	1095	CysSerSerLeuAlaIleArgGluMetGlnIleLysThrThrMetArgTyrHisLeu	1114
QY	29079	ACACCACTAGATGGCAATCATTAAAGTCAGGAACCAACAGGTCCTGGAGAGATGT	29020
Db	1115	ThrProValArgMetAlaIleIleLysLysSerGlyAsnAsnArgCysTrpArgGlyCys	1134
QY	29019	GGAAATAGGAACACTTTTACACTGTGTGGGACTGTGAACACTTCAACCATTTGG	28960
Db	1135	GlyGluIleGlyThrLeuLeuHisCysTrpTrpAspCysLysLeuValGlnProLeuTrp	1154
QY	28959	AAGTCAGTGGGTATTCCTCAGGGATCTAGAACTAGAATACCATTTGACCCAGCCATC	28900
Db	1155	LysSerValTrpArgPheLeuArgAspLeuGluLeuGluIleProPheAspProAlaIle	1174
QY	28899	CCATTACTGGGTATATACCAAGGATTAATAATCATGCTGCTATATAAGACACACACACA	28840
Db	1175	ProLeuLeuGlyIleTyrProGluAspTyrLysSerCysCysTrpLysAspThrCysThr	1194
QY	28839	TGTATGTTTATAGCAGCACTATTACAATAGCAAGACTTGGAAACCAACCTAAATGTCCA	28780
Db	1195	ArgMetPheIleAlaLeuPheThrIleAlaLysThrTrpAsnGlnProLysCysPro	1214
QY	28779	ACAACGATAGCTGGATTAAGAAATGTGGCACATATACCATGGAATCTATGACGCC	28720
Db	1215	ThrMetIleAspTrpIleLysLysMetTrpHisIleTyrThrMetGluTyrTyrAlaAla	1234
QY	28719	ATAAAAATGATGAGTTCGTCTCTTCTAGGGACATGATGAGTGAACCATCATTT	28660
Db	1235	IleLysAsnAspGluPheIleSerPheValGlyThrTrpMetLysLeuGluThrIleIle	1254
QY	28659	CTCAGCAAACTATCACAAGGCAAAAAACCAACCCGATGTTCTCTACTCATAGTGGG	28600
Db	1255	LeuSerLysLeuSerGlnGluGlnLysThrLysHisArgIlePheSerLeuIleGly	1274
QY	28599	AAT 28597	
Db	1275	Asn 1275	
RESULT 12			
line-1 protein ORF2 - human			
C:Species: Homo sapiens (man)			
C:Date: 03-Nov-1988 #sequence_revision 03-Nov-1988 #text_change 30-Sep-1993			
C:Accession: B28096			
R:Skowronski, J.; Fanning, T.G.; Singer, M.F.			
Mol. Cell. Biol. 8, 1385-1397, 1988			
A:Title: Unit-length line-1 transcripts in human teratocarcinoma cells.			
A:Reference number: A28096; MUID:88246405; PMID:2454389			
A:Accession: B28096			
A:Status: preliminary; not compared with conceptual translation			
A:Molecule type: mRNA			
A:Residues: 1-1275 <SKO>			
C:Superfamily: pol polyprotein			
Alignment Scores:			
Pred. No.: 1,27e-165 Length: 1275			
Score: 2196.00 Matches: 405			
Percent Similarity: 94.10% Conservative: 10			
Best Local Similarity: 91.84% Mismatches: 26			
Query Match: 4.19% Indels: 0			
DB: 2 Gaps: 0			
US-10-083-853B-2 (1-29921) x B28096 (1-1275)			
QY	29919	ACTTTCTCACAGAAATTCGAAAAAATACTTTAAAGTTTCATGTGAACCAAAAAAGAGCC	29860
Db	835	ThrPhePheThrGluLeuGluLysThrThrLeuLysPheIleTrpAsnGlnLysArgAla	854
QY	29859	CACATCGCCAGTCAATCCTTAAGCCAAAGAACAAAGCTGAGGCACTCATCTACCTGAC	29800
Db	855	ArgIleAlaLysSerIleLeuSerGlnLysAsnLysAlaGlyIleThrLeuProAsp	874
QY	29799	TTCAAACTACACTACAAGGCTACAGTAACCAACCAACATGGTACTGGTACCAAAACAGA	29740
Db	875	PhelYsLeuYrTyrLysAlaThrValThrLysThrAlaTrpTyrTrpTyrGlnAsnArg	894
QY	29739	GATATCAACCAATGGAACAGACAGCCCTCAGAAATAATGCCGATATCTACAACATAT	29680
Db	895	AspIleAspGlnTrpAsnArgThrGluProSerGluIleThr***LeuThrTyrAsnTyr	914
QY	29679	CTGATCTTTGACAAACCTGAGAAACCAACCAATGGGAAAGGATTCCTATTATAATAA	29620
Db	915	LeuIlePheAspLysProGluLysAsnLysGlnTrpGlyLysAspSerLeuPheAsnLys	934
QY	29619	TGGTGCTGAGAAACTGGCTAGCCCTATGTAGAAAGCTGAAACCTGGATCCCTCTTACA	29560
Db	935		

|||||MetSerLysThrProLysAlaMetAlaThrLysAlaLysIleAspLysTyrAspLeuIle 1014
QY 29379 AAACATAAGAGCTCTGCACAGCAAAAGAACTACCATCAGAGTGAACAGCAACTTACA 29320
|||||LysLeuLysSerPheCysThrAlaLysGluThrIleArgValAsnArgGlnProThr 1034
Db 1015 LysLeuLysSerPheCysThrAlaLysGluThrIleArgValAsnArgGlnProThr 1034
QY 29319 GAATGGGAGAAAATTTTGC AACCTACTCATTTGACAAAAGGCTTAATATCCAGAACTTAC 29260
|||||ThrIrpGluThrIlePheThrThrTyrSerSerAspLysGlyLeuIleSerArgIleTyr 1054
Db 1035 ThrIrpGluThrIlePheThrThrTyrSerSerAspLysGlyLeuIleSerArgIleTyr 1054
QY 29259 AATGAATCAACAAATTTTACAGAAAACAAAGAACCCCATCAAAAGTGGGTGAAG 29200
|||||AsnGluLeuLysGlnIleTyrLysLysThrAsnAsnProIleLysLysTyrAlaLys 1074
Db 1055 AsnGluLeuLysGlnIleTyrLysLysThrAsnAsnProIleLysLysTyrAlaLys 1074
QY 29199 GATATGAACACACTTCTCAAAAGAGGCAATTTATGCAGCAAAACACATCAAAAAA 29140
|||||AspMetAsnArgHisPheSerLysGluAspIleTyrAlaLysLysHisMetLysLys 1094
Db 1075 AspMetAsnArgHisPheSerLysGluAspIleTyrAlaLysLysHisMetLysLys 1094
QY 29139 TACTCATCATCAGTGGCCATCAGAGAAATGCAAAATCAAAACCAATGAGATACCATCTC 29080
|||||CysSerSerSerLeuAlaIleArgGluMetGlnIleLysThrThrMetArgTyrHisLeu 1114
Db 1095 CysSerSerSerLeuAlaIleArgGluMetGlnIleLysThrThrMetArgTyrHisLeu 1114
QY 29079 ACACAGTTAGATGGCAATCATTAAGTTCAGGAACAAACAGGTGCTGGAGAGAATGT 29020
|||||ThrProValArgMetAlaIleLysLysSerGlyAsnAsnArgCysTyrArgGlyCys 1134
Db 1115 ThrProValArgMetAlaIleLysLysSerGlyAsnAsnArgCysTyrArgGlyCys 1134
QY 29019 GGAGAAATAGAACACTTTTACACTGTTGGTGGGACTGTGAACCTAGTTCACCATTTGGG 28960
|||||GlyGluIleGlyThrLeuLeuHisCysTyrTrpAspCysLysLeuValGlnProLeuTyr 1154
Db 1135 GlyGluIleGlyThrLeuLeuHisCysTyrTrpAspCysLysLeuValGlnProLeuTyr 1154
QY 28959 AAGTCAGTGTGGTGAATTCCTCAGGGATCTAGAACTAGAAATPACCATTTGACCCAGCCATC 28900
|||||LysSerValTyrPheLeuArgAspLeuGluLeuGluIleProPheAspProAlaIle 1174
Db 1155 LysSerValTyrPheLeuArgAspLeuGluLeuGluIleProPheAspProAlaIle 1174
QY 28899 CCATTACTGGTATATACCCAAAGGATATATAATCATCTGCTCTATAAGACACAAAGCACA 28840
|||||ProLeuLeuGlyIleTyrProAsnAspTyrLysSerCysCysTyrLysAspThrCysThr 1194
Db 1175 ProLeuLeuGlyIleTyrProAsnAspTyrLysSerCysCysTyrLysAspThrCysThr 1194
QY 28839 TGTATGTTATAGCAGCACTATTCCACATAGCAAGACTTGGAAACCACTTAATGTGCCA 28780
|||||ACAAGTATAGTGGATTAAGAAAATGGCACATATACACCATGGAATCTATGCAGCC 28720
Db 1195 ArgMetPheIleAlaLeuPheThrIleAlaLysThrTrpAsnGlnProLysCysPro 1214
QY 28779 ACAAGTATAGTGGATTAAGAAAATGGCACATATACACCATGGAATCTATGCAGCC 28720
|||||ThrMetIleAspTyrIleLysLysMetTrpHisIleTyrThrMetGluTyrTyrAlaAla 1234
Db 1215 ThrMetIleAspTyrIleLysLysMetTrpHisIleTyrThrMetGluTyrTyrAlaAla 1234
QY 28719 ATAAAAATGATGATGCTGCTCTTTGTAGGGACATGGATGAAGCTGGAAACCATCATTT 28660
|||||IleLysAsnAspGluPheMetSerPheValGlyThrTrpMetLysLeuGluThrIleIle 1254
Db 1235 IleLysAsnAspGluPheMetSerPheValGlyThrTrpMetLysLeuGluThrIleIle 1254
QY 28659 CTCAGCAACATPACACAGGACAAAACCAACCAACCGCATGTTCTCACTCATAGGTGGG 28600
|||||LeuSerLysLeuSerGlnGluLysThrLysHisArgIlePheSerLeuIleGlyGly 1274
Db 1255 LeuSerLysLeuSerGlnGluLysThrLysHisArgIlePheSerLeuIleGlyGly 1274
QY 28599 AAT 28597
|||||
Db 1275 Asn 1275

RESULT 13
S23650

retrovirus-related hypothetical protein II - human retrotransposon LINE-1

C:Species: Homo sapiens (man)

C:Date: 22-Nov-1993 #sequence_revision 01-Nov-1996 #text_change 08-Jan-1999

C:Accession: S23650

R:Hohjoh, H.; Minakami, R.; Sakaki, Y.

Nucleic Acids Res. 18, 4099-4104, 1990

A:Title: Selective cloning and sequence analysis of the human L1 (LINE-1) sequences which

A:Reference number: S23649; MUID:90332398; PMID:2165587

A:Accession: S23650

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-712 <HOH>

A:Cross-references: EMBL:X52235
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1990
C:Genetics:
A:Mobile element: LINE-1
A:Start codon: GTG
C:Superfamily: pol polyprotein

Alignment Scores:

Pred. No.: 1,14e-164 Length: 712
Score: 2183.00 Matches: 403
Percent Similarity: 93.88% Conservative: 11
Best Local Similarity: 91.38% Mismatches: 27
Query Match: 4.17% Indels: 0
DB: 2 Gaps: 0

US-10-083-853B-2 (1-29921) x S23650 (1-712)

QY 29919 ACTTCTTCACAGAAATTTGAAAAAACTACTTTAAAGTTTCATGTGGAACCAAAAAAGAGCC 29860
|||||ThrPhePheThrGluLeuGluLysThrThrLeuLysPheIleTyrAsnGlnLysArgAla 291
Db 272 ThrPhePheThrGluLeuGluLysThrThrLeuLysPheIleTyrAsnGlnLysArgAla 291
QY 29859 CACATCGCCCAAGTCAATCTTAAGCCAAAAGAAAGCTGGAGGCATCATCTACCTGAC 29800
|||||ArgIleAlaLysAlaIleArgSerGlnLysAsnLysSerGlyIleThrLeuProAsp 311
Db 292 ArgIleAlaLysAlaIleArgSerGlnLysAsnLysSerGlyIleThrLeuProAsp 311
QY 29799 TTCAAACTACACTACAAGGCTACAGTAACCAAAAAACATCGTACTGTGTCACCAAAACAGA 29740
|||||PhelysLeuTyrTyrLysAlaThrValThrLysThrAlaTyrTyrTyrGlnAsnArg 331
Db 312 PhelysLeuTyrTyrLysAlaThrValThrLysThrAlaTyrTyrTyrGlnAsnArg 331
QY 29739 GATATCAACCAATGGAACAGAACAGAGCCCTCAGAAATAATGTCGCATATCTACCACTAT 29680
|||||AspIleAspGlnTyrPasnArgThrGluProSerGluIleThrProHisIleTyrAsnTyr 351
Db 332 AspIleAspGlnTyrPasnArgThrGluProSerGluIleThrProHisIleTyrAsnTyr 351
QY 29679 CTGATCTTTGACAAACCTGACAAAACAAGCAATGGGAAAGGATTCCTCTTTAAATAAA 29620
|||||LeuIlePheAspLysProGluLysAsnGluGlnTyrGlyLysAspSerLeuPheAsnLys 371
Db 352 LeuIlePheAspLysProGluLysAsnGluGlnTyrGlyLysAspSerLeuPheAsnLys 371
QY 29619 TGGTCTGAGAAAAAATCGCTAGCCCTATGTAGAAGCTGAAACCTGATCCCTCTTCCATACA 29560
|||||TrpCysTyrTrpGluAsnTyrLeuAlaIleCysArgLysLeuLysLeuAspProPheLeuThr 391
Db 372 TrpCysTyrTrpGluAsnTyrLeuAlaIleCysArgLysLeuLysLeuAspProPheLeuThr 391
QY 29559 CCTATACAAAATTAATTCAGATGGATTAAAGACTTAATGTTAGACCTAAACCAATA 29500
|||||ProTyrThrLysIleAsnSerArgTyrIleLysAspLeuIleValArgProLysThrIle 411
Db 392 ProTyrThrLysIleAsnSerArgTyrIleLysAspLeuIleValArgProLysThrIle 411
QY 29499 AAATCCCTAGAGAAAAACCCAGCAATACCATTCAGGACATAGGCATGGCAAGGACTTC 29440
|||||LysThrLeuGluGluAsnLeuGlyIleThrIleGlnAspIleGlyMetGlyLysAspPhe 431
Db 412 LysThrLeuGluGluAsnLeuGlyIleThrIleGlnAspIleGlyMetGlyLysAspPhe 431
QY 29439 ATGTCTAAACACCAAAAGCAATGGCAACAAAGCCAAAATTTGACAAATGGATCGATCAATT 29380
|||||MetSerLysThrProLysAlaMetAlaThrLysAlaLysIleAspLysTyrAspLeuIle 451
Db 432 MetSerLysThrProLysAlaMetAlaThrLysAlaLysIleAspLysTyrAspLeuIle 451
QY 29379 AAATAAGAGCTTCTGCACAGAAAAGAACTACCATCAGAGTGAACAGCAACTTACA 29320
|||||LysLeuLysSerPheCysThrAlaLysGluThrThrIleArgValAsnArgGlnProThr 471
Db 452 LysLeuLysSerPheCysThrAlaLysGluThrThrIleArgValAsnArgGlnProThr 471
QY 29319 GAATGGGAGAAAATTTTGGCAACCTACTCTTTGACAAAGGCTTAATATCCAGAACTTAC 29260
|||||LysIrpGluLysThrPheAlaThrTyrSerSerAspLysGlyLeuIleSerArgIleTyr 491
Db 472 LysIrpGluLysThrPheAlaThrTyrSerSerAspLysGlyLeuIleSerArgIleTyr 491
QY 29259 AATGAATCAACAAATTTTACAGAAAACAAAGAAACCCCATCAAAAGTGGGTGAAG 29200
|||||AsnGluLeuLysGlnIleTyrLysLysLysThrAsnAsnProIleLysLysTyrAlaLys 511
Db 492 AsnGluLeuLysGlnIleTyrLysLysLysThrAsnAsnProIleLysLysTyrAlaLys 511
QY 29199 GATATGAACACACTTCTCAAAAGAGGCAATTTATGCAGCAAAACACATCAAAAAA 29140
|||||AspMetAsnArgHisPheSerLysGluAspIleTyrAlaLysLysHisMetLysLys 531
Db 512 AspMetAsnArgHisPheSerLysGluAspIleTyrAlaLysLysHisMetLysLys 531
QY 29139 TACTCATCATCAGTGGCCATCAGAGAAATGCAAAATCAAAACCAATGAGATACCATCTC 29080
|||||CysSerProSerLeuAlaIleArgGluMetGlnIleLysThrThrMetArgTyrHisLeu 551
Db 532 CysSerProSerLeuAlaIleArgGluMetGlnIleLysThrThrMetArgTyrHisLeu 551

```
QY 29079 ACACAGTGAATGGAATCATATAAAGTCAGGAAACACAGGCTGCTGGAGAGATGT 29020
Db |||
Db 552 ThrProValArgMetAlaIleLeuLysLysSerGlyAsnAsnArgCysTrpArgGlyCys 571
QY 29019 GGAGAAATAGAACACTTTTACACTGTTGGTGGAGCTGCTCAACTAGTTCACCACTGTGG 28960
Db |||
Db 572 GlyGluIleGlyThrLeuLeuHisCysTrpTrpAspCysLysLeuValGlnProLeuTrp 591
QY 28959 AAGTCAGTGTGGTGATTCCTCAGGAGTCTAGAACTAGAAATACCATTTGACCCAGCCATC 28900
Db |||
Db 592 LysSerValTrpArgPheLeuArgAspLeuGluProGluIleProPheAspProAlaIle 611
QY 28899 CCATTACTGGGTATATACCAAGGATTATAAATCATGCTGCTATAAAGACACAAACACA 28840
Db |||
Db 612 ProLeuLeuGlyIleYrProLysAspSerLysSerCysCysTrpLysAspThrCysThr 631
QY 28839 TGTATGTTTATACAGCAGCTATTTCACAATAGCAAGACACTTGGAAACCACTTAATGTCCA 28780
Db |||
Db 632 ArgMetPheIleAlaLeuPheThrIleAlaLysThrTrpAsnGlnProLysCysPro 651
QY 28779 ACAACGATAGACTGGATTAGAAATGTGGCACATATACACCATGGAATACTATGCAGCC 28720
Db |||
Db 652 ThrMetIleAspTrpIleLysLysMetTrpHisIleYrThrMetGluYrTrpAlaAla 671
QY 28719 ATAAAAATGATGAGTTGCTGCTCTTTGAGGACATGGATGAAGCTGGAACCATCAT 28660
Db |||
Db 672 IleLysAsnAspGluPheValSerPheValGlyThrTrpMetLysLeuGluIleIle 691
QY 28659 CTGAGCAAACTATCACAGGACAAAAACCAACACCGGATGTTCTCACTCATAGTGGG 28600
Db |||
Db 692 LeuSerLysLeuSerGlnGluGlnLysThrLysHisCysIlePheSerLeuIleGlyGly 711
QY 28599 AAT 28597
Db |||
Db 712 Asn 712

RESULT 14
JU0033
hypothetical L1 protein (third intron of gene TS) - human
C:Species: Homo sapiens (man)
C:Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 13-Sep-1996
C:Accession: JU0033
R:Horie, N.; Naibantoglu, J.; Kaneda, S.; Ayusawa, D.; Seno, T.; Takeishi, K.
J. Biochem. 106, 1-4, 1989
A:Title: Identification and characterization of an L1 family sequence with a very long
A:Reference number: JU0033; PMID:89380111; PMID:2476429
A:Accession: JU0033
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-562 <HOR>
C:Superfamily: pol polyprotein

Alignment Scores:
Pred. No.: 3,49e-164 Length: 562
Score: 2176.50 Matches: 449
Percent Similarity: 85.82% Conservative: 35
Best Local Similarity: 79.61% Mismatches: 74
Query Match: 4.16% Indels: 10
DB: 2 Gaps: 2

US-10-083-853b-2 (1-29921) x JU0033 (1-562)

QY 5429 CAGCATCATCTCGATACCAAAACCTAGCAGTGACACACAAAAAGAGAAATTTTCAGGCC 5488
Db |||
Db 2 GlnHisHisProAspThrLysAlaGlyGlnArgHis-ThrLysLysGluAsnPheArgPr 21
QY 5489 CATATCCCTGATGAACATTGATGTGAAATCTCAATAAATCTCAATAAATCTGCGCAAAACCAATCCA 5548
Db |||
Db 21 oileSerLeuMetAsnIleAspAlaLysIleLeuAsnLysIleLeuAlaAsnArgIleG 41
QY 5549 GCAGCACATCAAAAGCTTATCTACCATGATCAAGTTGGCGTCATCCCTGGGATGCAAGG 5608
Db |||
```

```
Db 41 nGlnHisIleLysLysLeuIleHisAspHisValGlyPheIleProArgMetGlnG 61
QY 5609 CTGGTTCAAAATATGCAAAATCAATAATGTAGGCCATCATATAACAGAACCAATGACAA 5668
Db |||
Db 61 YrTrpPheAsnIleArgLysSerIleAsnValIleGlnHisIleAsnArgSerLysAspLys 81
QY 5669 AAACACCATGATTATCTCAATAGATGCAGAAAAAGCCTTTGTCTCAAAATTCACACAGCCCTT 5728
Db |||
Db 81 sasnHisMetIleIleSerIleAspAlaGluLysAlaPheAspLysIleGlnGlnHisI 101
QY 5729 CATGCTAAAAATCTCAGTAACTAGTATCGATGGAATGTATCTCAAAATATAATAGAGC 5788
Db |||
Db 101 eMetLeuLysThrLeuAsnLysLeuGlyIleAspGlyThrYrPheLysIleIleArgAl 121
QY 5789 TATTTATAC-AAACCCACACAGCCAAATATCATCTAGTAAATGGGCAAAAACTTGAAGCATTC 5847
Db |||
Db 121 alleYrAspLysAlaThrAlaAsnIleIleValAsnGlnLysLeuGluAlaPhePr 141
QY 5848 TTTGAGAACTGGCACAGACAGGATGCCCTCTCTCACCACTCCTATTTCAGAACTACTATT 5907
Db |||
Db 141 oLeuLysThrGlyThrArgGlnGlyCysProLeuSerProLeuLeuPheAsnIleValle 161
QY 5908 GGAACTTCGGCCAGGCAATCAGGCAATAGAAAGAAATAAAGGGTATTCAATAGANAAG 5967
Db |||
Db 161 uGluValLeuAlaArgAlaIleArgGlnLysGluIleLysGlyIleGlnLeuGlyLy 181
QY 5968 AGAGGAAGTCATATTCTCTCTGTCAGATGACATGTTGTATATTATAGAAAACCCCAT 6027
Db |||
Db 181 sGluGluValLysLeuSerLeuPheAlaAspMetIleValYrLeuGluAsnProIl 201
QY 6028 CGTCTCAGGCCAAAACTCCTTAAGCTGATAGCAACTTCAGCAAAAGTCTCAGGACACAA 6087
Db |||
Db 201 eValSerAlaGlnAsnLeuLysLeuIleSerSerPheSerLysValSerGlyYrLy 221
QY 6088 AATCAATGTCRAAAATCACAACCATTCATTATACGCCCAATATAGACAAACAGAGAGCCA 6147
Db |||
Db 221 sIleAsnLeuGlnLysSerGlnAlaPheLeuYrThrAsnAsnArgGlnThrGluSerGl 241
QY 6148 AATCATGAGTGAACCTCTCATTCACAATTCGTACAAAGAGAATAAATACTAGGAATACA 6207
Db |||
Db 241 nIleMetSerGluLeuProLeuThrIleAlaSerLysArgIleLysYrLeuGlyIleH 261
QY 6208 ACTTACAGGGACACCTAGGAACCTTCTCAAGGAGAACTCAACACCACTGATCAAGGAAT 6267
Db |||
Db 261 sLeuLysArgAspValLysAspLeuPheLysGluAsnYrLysProLeuLeuAsnGluIl 281
QY 6268 AAGAGAGGACACAAACAAATGGAACCATTCATGCTCACAGATAGTAGAATCAT--- 6324
Db |||
Db 281 eLysGluAspThrLysLysTrpLysThrIleProCysSerTrpValGlyArgIleAsnIl 301
QY 6325 ---GAAAAATGCCATCTGCCAAAGTAAATATATAGATTCACTGCTACCCCTCAAGCT 6380
Db |||
Db 301 eValLysMetAlaIleLeuProLysValIleYrArgPheAsnAlaIleProIleLysLe 321
QY 6381 ACATTGACTTCTTTCACAGAATTTGAAAAACAACTTTAAATTTTCATATGGAACCAAAA 6440
Db |||
Db 321 uProMetThrPhePheThrGluLeuGluLysThrThrLeuLysPheIleTrpAsnGln-L 341
QY 6441 AAGAGCCACACAGAGCCAAAGCAATCTTAAGCAAAAGAAAGCAAGCTGGAGGTATCATGC 6500
Db |||
Db 341 ysArgAlaArgIleAlaLysSerIleLeuSerGlnLysAsnLysAlaGlyIleThrL 361
QY 6501 TACCTGACTTAAACTATATAGGCTACAGTACCAAAACTGCATGGTACTGTGCTACC 6560
Db |||
Db 361 euProAspPheLysLeuYrYrLysAlaThrValThrLysThrValTrpYrTrpTyG 381
QY 6561 AAACAGATATATAGACCAATGGACACAGACAGACCTCAGAAATTCAGACT-GCAATCT 6619
Db |||
Db 381 InAsnArgHisIleAspGlnTrpAsnArgThrGluProSerGluIleThrProHisIle 401
QY 6620 ACATCCATCTGATCTTTGACAAACCTTGACAAACCAAGCAATCGAAAAAGGATTCCTCAT 6679
Db |||
Db 401 yrasnYrLeuIlePheAspLysProGluLysAsnLysIntrpGlyLysAspSerLeuc 421
```

Qy	6680	TTATAAATGGTGTGGAAAGCTAGCCATATGCAAGAGCTGAAAGCTGATCCCT	6739	Db	894	AspLeuAspGlnTrpAsnArgThrGluProSerGluLeuMetProHisIleTyrAsnTyr	913
Db	421	ysAsnLysTyrCysTrpGluAsnTrpLeuAlaIleCysLysLysLeuLysLeuAspProp	441	Qy	29679	CTGATCTTTACAAACCTGAGAAAACAAGCAATGGGGAAGGATTCCTTATTAAATAA	29620
Qy	6740	TCCTTACACCTTATACAAAAGTTAACTCAAGATGAATTAAGACTTAAATATAGACATA	6799	Db	914	LeuIlePheAspLysProGluLysAsnLysGlnTrpGlyLysAspSerLeuPheAsnLys	933
Db	441	heLeuThrProTyrThrLysIleAsnSerArgTrpIleLysAspLeuAsnValArgProL	461	Qy	29619	TGGTCTGAGAAAACCTAGCCCTATGTAGAAAGCTGAAAGCTGGATCCCTCTCTTACA	29560
Qy	6800	AAACCATAAACCCCA-GAAGAAACCTAGGCAATACCATTCAGGATATGACATGGCA	6858	Db	934	TrpCysTrpGluAsnTrpLeuAlaIleCysArgLysLeuLysLeuAspProPheLeuThr	953
Db	461	ysThrIleLysThrLeuGluGluAsnLeuGlyIleThrIleGlnAspIleGlyMetGlyL	481	Qy	29559	CCTTATACAAAATAATTAATCAAGATGATTAAGACTTAAATGTAGACCTTAAACCAT	29500
Qy	6859	AAGACTTCATGATAGAGCTTCTGACAGCAAAAGGCAAAAGGCAAAATAGCAAGTGG	6918	Db	954	ProTyrThrLysIleAsnSerArgTrpIleLysAspLeuAsnValArgProLysThrIle	973
Db	481	ysAspPheMet---PrometProLysAlaMetAlaThrLysAlaLysIleAspLysTrpA	500	Qy	29499	AAATCCCTAGAAAGAAACCCAGGCAATACCATTCAGGACATAGGCATGGCAAGACTTC	29440
Qy	6919	ATCTGATTAACATATAGAGCTTCTGACAGCAAAAGGCAAAAGGCAAAATAGCAAG	6978	Db	974	LysThrLeuGluLysAsnLeuGlyAsnThrIleGlnAspIleGlyMetGlyLysAspPhe	993
Db	500	spLeuIleLysLeuLysSerPheCysThrAla-LysGluThrThrIleArgValAsnArg	519	Qy	29439	ATGTCTAAAACACCAAAAGCAATGGCAACAAAAGCCAAAATTTGCAAAATGGATCTAAT	29380
Qy	6979	CAACCTACAGAAATGGGAGAAAATTTTCAATCTATCGATCTGACAAAGGCTAATATCCA	7038	Db	994	MetThrLysThrProLysAlaMetAlaThrLysValLysIleAspArgTrpAspLeuIle	1013
Db	520	GlnProThrLysTrpGluAsnIlePheAlaThrTyrSerSerAspLysGlyLeuIleSer	539	Qy	29379	AACTAAAGAGCTTCTGCACAGCAAAAGAACTACCATCAGATGAGAGCAACTTACA	29320
Qy	7039	GAGATCAGCAAGAACTTAAACAAATTTACAAGAAAAAACAACCCGCTCAAAATATGG	7098	Db	1014	LysLeuLysSerPheCysThrAlaLysGluThrThrIleArgValAsnArgGlnProThr	1033
Db	540	ArgIleTyrAsnGluLeuLysGlnIleTyrLysLysLysThrThrProSerLysSerGly	559	Qy	29319	GAATGGAGAAAATTTTGCACCTACTCATTTGACAAAGGGCTAATATCAGAACTCAC	29260
Qy	7099	CAARGG 7104		Db	1034	LysTrpGluLysIlePheAlaIleTyrSerSerAspLysGlyLeuIleSerArgIleTyr	1053
Db	560	ArgArg 561		Qy	29259	AATCAACTCAACAAATTTTACAAGAAAAAACAAGAAACCCCATCAAAAAGTGGTGAAG	29200
RESULT 15				Db	1054	AsnGluLeuLysGlnIleTyrLysLysLysThrAsnAsnProIleLysLysTrpAlaLys	1073
GNHULL				Qy	29199	GATATGACAGACACTTCTCAAAGAGGCGATTTATGCAGCCAAAACACATCAAAAAA	29140
retrovirus-related reverse transcriptase pseudogene - human							
C:Species: Homo sapiens (man)							
C>Date: 31-Mar-1998 #sequence_revision 04-Jan-1996 #text_change 14-May-1999							
C:Accession: A25313							
R:Hattori, M.; Kuhara, S.; Takenaka, O.; Sakaki, Y.							
Nature 321, 625-628, 1986							
A:Title: L1 family of repetitive DNA sequences in primates may be derived from a sequence							
A:Reference number: A93381; MUID:86230917; PMID:2423883							
A:Accession: A25313							
A>Status: conceptual translation of pseudogene							
A:Molecule type: DNA							
A:Residues: 1-1259 <HAT>							
A:Note: this sequence was constructed from an alignment of published and unpublished seq							
C:Keywords: reverse transcriptase; pseudogene							
Alignment Scores:							
Pred. No.:	1.44e-161	Length:	1259				
Score:	2145.00	Matches:	394				
Percent Similarity:	95.07%	Conservative:	11				
Best Local Similarity:	92.49%	Mismatches:	21				
Query Match:	4.09%	Indels:	0				
DB:	4	Gaps:	0				
US-10-083-853B-2 (1-29921) x GNHULL (1-1259)							
Qy	29919	ACTTCTTCACAGAAATGGAAAAAAGCTAAAGTTCATGTGGAACCAAAAAAGAGCC	29860	Db	834	ThrPhePheThrGluLeuGluLysThrThrLeuLysPheIleThrPheAsnGlnLysArgAla	853
Db	29859	CACATCCCAAGTCAATCCCTAAGCCAAAGAAAGCAAGCTGGAGGCATCATCTACCTGAC	29800	Qy	29859	CACATCCCAAGTCAATCCCTAAGCCAAAGAAAGCAAGCTGGAGGCATCATCTACCTGAC	29800
Db	854	HisIleAlaLysSerThrLeuSerGlnLysAsnLysAlaGlyGlyIleThrLeuProAsp	873	Qy	29799	TTCAAACTACACTACAAGGCTACAGTAACCAAAACACATGGTACTGGTACCAAAACAGA	29740
Qy	29799	TTCAAACTACACTACAAGGCTACAGTAACCAAAACACATGGTACTGGTACCAAAACAGA	29740	Db	874	PhelysLeuTyrLysAlaThrValThrLysThrAlaTrpTyrTrpTyrGlnAsnArg	893
Db	29739	GATATCAACCAATGGAAACAGACAGAGCCCTCAGAAATAATGCGCATATCTACAACTAT	29680	Qy	28659	CTCAGCAAACTATCACAA 28642	

Db 1254 LeuSerLysLeuSerGln 1259

Search completed: July 4, 2003, 22:37:08
Job time : 785 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 4, 2003, 20:41:28 ; Search time 515.5 Seconds

(without alignments)
13348.993 Million cell updates/sec

Title: US-10-083-853b-2

Refseq scores: 52328

Sequence: 1 gatatgtgaagaagctca.....caattctgtgaagaagtaa 29921

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 440863 seqs, 114992915 residues

Total number of hits satisfying chosen parameters: 881726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10083853/runat_03072003_093613_8176/app_query.fasta_1.30087
-DB=PublishedApplications_AA -QMT=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USPR=US10083853 @CGN_1_1_192_8runat_03072003_093613_8176
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database 8 Published Applications AA

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/FCI_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5169	9.9	1275	9 US-10-025-201-3	Sequence 3, Appli
2	4081.5	7.8	1010	9 US-10-050-882-80	Sequence 80, Appli
3	3770.5	7.2	956	9 US-10-000-256A-153	Sequence 153, App
4	3628.5	6.9	917	9 US-10-001-835-189	Sequence 189, App

5	3535.5	6.8	940	9 US-10-082-830-235	Sequence 235, App
6	2215	4.2	1275	9 US-10-025-201-3	Sequence 3, Appli
7	2199	4.2	1010	9 US-10-050-882-80	Sequence 80, Appli
8	2187	4.2	956	9 US-10-000-256A-153	Sequence 153, App
9	1467	2.8	917	9 US-10-001-835-189	Sequence 189, App
10	1351	2.6	940	9 US-10-082-830-235	Sequence 235, App
11	1090.5	2.1	338	9 US-10-025-201-2	Sequence 2, Appli
12	806	1.5	317	9 US-10-083-853-2	Sequence 2, Appli
13	767.5	1.5	341	9 US-10-001-876-169	Sequence 169, App
14	708	1.4	492	9 US-10-007-280A-212	Sequence 212, App
15	676	1.3	190	9 US-10-074-475-250	Sequence 250, App
16	630	1.2	492	9 US-10-007-280A-212	Sequence 212, App
17	590.5	1.1	315	12 US-10-001-843-125	Sequence 125, App
18	580	1.1	192	9 US-10-001-857-119	Sequence 119, App
19	573.5	1.1	317	9 US-10-083-853-2	Sequence 2, Appli
20	548.5	1.0	197	9 US-10-157-031-367	Sequence 367, App
21	533.5	1.0	192	9 US-10-001-857-119	Sequence 119, App
22	533	1.0	315	12 US-10-001-843-125	Sequence 125, App
23	481.5	0.9	212	10 US-09-864-761-46654	Sequence 46654, A
24	460.5	0.9	230	9 US-10-000-256A-143	Sequence 143, App
25	434	0.8	113	9 US-10-012-542-399	Sequence 399, App
26	431.5	0.8	316	12 US-10-001-870-132	Sequence 132, App
27	429.5	0.8	121	9 US-10-106-698-4424	Sequence 4424, Ap
28	429	0.8	167	10 US-09-864-761-34496	Sequence 34496, A
29	414.5	0.8	101	9 US-10-012-542-400	Sequence 400, App
30	414	0.8	121	9 US-10-106-698-4424	Sequence 4424, Ap
31	409	0.8	322	9 US-10-086-623-6	Sequence 6, Appli
32	409	0.8	322	9 US-10-260-539-6	Sequence 6, Appli
33	409	0.8	364	9 US-10-028-072-186	Sequence 186, App
34	409	0.8	364	9 US-10-121-049-186	Sequence 186, App
35	409	0.8	364	9 US-10-123-904-186	Sequence 186, App
36	409	0.8	364	9 US-10-140-470-186	Sequence 186, App
37	409	0.8	364	9 US-10-175-746-186	Sequence 186, App
38	409	0.8	364	9 US-10-176-918-186	Sequence 186, App
39	409	0.8	364	9 US-10-176-921-186	Sequence 186, App
40	409	0.8	364	9 US-10-137-865-186	Sequence 186, App
41	409	0.8	364	9 US-10-140-474-186	Sequence 186, App
42	409	0.8	364	9 US-10-142-431-186	Sequence 186, App
43	409	0.8	364	9 US-10-143-114-186	Sequence 186, App
44	409	0.8	364	9 US-10-140-002-186	Sequence 186, App
45	409	0.8	364	9 US-10-142-419-186	Sequence 186, App

ALIGNMENTS

RESULT 1
US-10-025-201-3
; Sequence 3, Application US/10025201
; Publication No. US20030003468A1
; GENERAL INFORMATION:
; APPLICANT: Crow, Mary K.
; TITLE OF INVENTION: MARKERS FOR DISEASE SUSCEPTIBILITY AND TARGETS FOR THERAPY
; FILE REFERENCE: 5983/2H567
; CURRENT APPLICATION NUMBER: US/10/025,201
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/256,673
; PRIOR FILING DATE: 2000-12-19
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 1275
; TYPE: PRT
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: GenBank Accession No. US20030003468A1 U09116
; DATABASE ENTRY DATE: 1995-02-02
; RELEVANT RESIDUES: (1)..(1275)
US-10-025-201-3

Alignment Scores:
Pred. No.: 0
Score: 5169.00
Percent Similarity: 78.90%
Length: 1275
Matches: 1054
Conservative: 79

Best Local Similarity: 73.40%		Mismatches: 137	
Query Match: 9.88%		Indels: 173	
DB:		Gaps: 7	
US-10-083-853B-2 (1-29921) x US-10-025-201-3 (1-1275)			
QY	3428	ATGGCAGGATCAAAATTCACACATAATATATAACCTTAAATGTAATGGCTAAATTC	3487
DB	1	MetThrGlySerAsnSerHisIleThrIleLeuThrLeuAsnIleAsnGlyLeuAsnSer	20
QY	3488	CCAATTAAGACACAGACTGGCAAAATTTGGATAAGAGTCAAGACCCATCAGTGTGCTGT	3547
DB	21	AlaIleLysArgHisArgLeuAlaSerTrpIleLysSerGlnAspProSerValCysCys	40
QY	3548	ATTGAGAGGCCATCTCACATGAAAGACACACATAGGCTCAAAATAAAGGGATGGAGG	3607
DB	41	IleGlnGluThrHisLeuMetCysArgAspThrHisArgLeuLysIleLysGlyTrpArg	60
QY	3608	AGATTTACCAAGTAATTTGGAACAAACAAAAAGCAGGGTTGCAATCCTAGTCTCT	3667
DB	61	LysIleTyrGlnAlaAsnGlyLysGln---LysLysAlaGlyValAlaIleLeuValSer	79
QY	3668	GATAAAGACAGCTTTAAACCAACAAGATCAAAAGACACAAAGAGGCCATTACATAATG	3727
DB	80	AspLysThrAspPheLysProThrLysIleLysArgAspLysGluGlyHisTyrIleMet	99
QY	3728	GTAAGGCATCAATGGACAAGAGAGCTAACTATCCTAAATATACATGACCCCAATACA	3787
DB	100	ValLysGlySerIleGlnGlnGluLeuThrIleLeuAsnIleTyrAlaProAsnThr	119
QY	3788	GGAGCACCCAGATTCATTAAGCAAGTCTTAGAGACCTTACAAAGAGACTTTGACTCCAC	3847
DB	120	GlyAlaProArgPheIleLysGlnValLeuSerAspLeuGlnArgAspLeuAspSerHis	139
QY	3848	ACAATAATAGTGGGAGTCTAAATAATAATAGACACTTTAACACCCCACTGCCAATATTA	3907
DB	140	ThrLeuIleMetGly-----AspPheAsnThrProLeuSerIleLeu	153
QY	3908	GGCAGATCAATGAGACAGAAATTAACAAGGATATCCAGGAGTTGAACAGCTCTGGAC	3967
DB	154	AspArgSerThrArgGlnLysValAsnLysAspThrGlnGluLeuAsnSerAlaLeuHis	173
QY	3968	CAAGCGGCTAATAGATATCTACAGAATCTCCCAACCCCAATCAACAGAAATATACACTC	4027
DB	174	GlnAlaAspLeuIleAspIleTyrArgThrLeuHisProLysSerThrGluTyrThrPhe	193
QY	4028	TTCTCAGCATCACATTAACCATTTTAAATTTACCATGTAATTTTAACTAAACACTC	4087
DB	194	PheSerAlaProHisHisThrTyrSerLysIleAspHisIleValGlySerLysAlaLeu	213
QY	4088	CTCAGCAATGCAAAAGAACAGAAATCTTAAACAAACAGTCTCTCAGACTACAGTCAATC	4147
DB	214	LeuSerLysCysLysArgThrGluIleIleThrAsnTyrLeuSerAspHisSerAlaIle	233
QY	4148	TATTTAGACTCAGAAATTAAGAAATCTACTCAAAATCACACACTACATGGAACACTGAAC	4207
DB	234	LysLeuGluLeuArgIleLysAsnLeuThrGlnSerArgSerThrThrTrpLysLeuAsn	253
QY	4208	ACCTGCTCCTGAAATGACTACTGGTAAATAACAAATGAAGGCAAAATAAAGATGTTTC	4267
DB	254	AsnLeuLeuLeuAsnAspTyrTrpValHisGlnGluMetLysAlaGluIleLysMetPhe	273
QY	4268	TTTGAACCAATGACAACAAAGACACATGTACCAGAAATCTCTGGGCGCATATTTAAAGCA	4327
DB	274	PheGluThrAsnGluAsnLysAspThrThrTyrGlnAsnLeuTrpAspAlaPheLysAla	293
QY	4328	GTGTGTAGAGGAAATTTATAGCACTAGATGCCTTACAGAGAAAGACGAGAAATATCTAAA	4387
DB	294	ValCysArgGlyLysPheIleAlaLeuAsnAlaTyrLysArgLysGlnGluArgSerLys	313
QY	4388	ATAGACACCTTAACATCAATTTAAAGAACTAGAGAGAAAGACGAAACAAATTCAAA	4447
DB	314	IleAspThrLeuThrSerGlnLeuLysGluLeuGlnGlnGlnGlnGlnGlnGlnGlnGln	333

QY	5521	TCATATAAATACTGGCAACCAATCCAGCAGCACATCAAAAGACTTATCTACCATGATC	5580	Db	903	GluproSerGluIleMetProHisIleTyrAsnTyrLeuIlePheAspLysProGluLys	922
Db	543	eAsnLysIleLeuAlaAsnArgIleGlnGlnHisIleLysLysLeuIleHisIleAsp	563	QY	6652	AACAAGCAATGGAAGGATTCCCTATTTAAATAAGTGTCTGGAAACACTGGCTAGCC	6711
QY	5581	AAGTTGGCGTCATCCCTGGGATGCAAGCTGGTTCAAAATATGCAATCAATAATGAT	5640	Db	923	AsnLysGlnTrpGlyLysAspSerLeuPheAsnLysTrpCysTrpGluAsnTrpLeuAla	942
Db	563	InValGlyPheIleProGlyMetGlnGlyTrpPheAsnIleArgLysSerIleAsnValI	583	QY	6712	ATATGCAAGAACTGAACTGGATCCCTTCCATCCCTTACACCTTATACAAAAGTTAACTCAAGA	6771
QY	5641	GCCATCATATAAAGCAACCAATGACAAAACACCATGATTATCTCAATAGATGCAAAA	5700	Db	943	IleCysArgLysLeuLysLysLeuAspLeuPheLeuThrProTyrThrLysIleAsnSerArg	962
Db	583	leGlnHisIleAsnArgAlaLysAspLysAsnHisMetIleIleSerIleAspAlaGluL	603	QY	6772	TGAATTAAGACTTAATATATAAGACATATAAACCACTATAAACCCCA-GAAGAAAACCTAGGC	6830
QY	5701	AGGCTTTGTCAAAATTCACAGCCCTTCATGCTAAATAATCTCAGTAAACTAGGTATCG	5760	Db	963	TrpIleLysAspLeuAsnValLysProLysThrIleLysThrLeuGluGluAsnLeuGly	982
Db	603	ysAlaPheAspLysIleGlnGlnProPheMetLeuLysThrLeuAsnLysLeuGlyIleA	623	QY	6831	AATACCATTCAGGATATGGACATGGCAAGACATTCATGACTATAACACCAACCAAGCAATG	6890
QY	5761	ATGCAATGTATCTCAAAATATAAGACTATTTATAC-AAACCCACAGCCCAATATCATAC	5819	Db	983	IleThrIleGlnAspIleGlyValGlyLysAspPheMetSerLysThrProLysAlaMet	1002
Db	623	spGlyThrTyrPheLysIleIleArgAlaIleTyrAspLysProThrAlaAsnIleIleL	643	QY	6891	GCAACAAAGCCAAATAGACAGTGGGATCTGATTAAACTATAGAGCTTCTGCACAGCA	6950
QY	5820	TGAATGGCAAAACTGGAAGCATTCCTTTTGAAACTGGCACAGCAAGGATGCCCTC	5879	Db	1003	AlaThrLysAspLysIleAspLysTrpAspLeuIleLysLeuLysSerPheCysThrAla	1022
Db	643	euAsnGlyGlnLysLeuGluAlaPheProLeuLysThrGlyThrArgGlnGlyCysProL	663	QY	6951	AAAAAACTGTCATCATCAGTGAACCAACCACTACAGATGGGAGAGAAATTTTGGCAAT	7010
QY	5880	TCTCACACCTCTATCAAGTACTATTGGAACTTCTGGCAGGCAATCAGGCAATAGA	5939	Db	1023	-LysGluThrThrIleArgValAsnArgGlnProThrThrTrpGluLysIlePheAlaTh	1042
Db	663	euSerProLeuLeuPheAsnIleValLeuGluValLeuAlaIleArgGlnGluL	683	QY	7011	CTATCGATCTCACAAAGCGCTAATATCCAGAGATCTACGAGAACTTAACCAAAATTTACAA	7070
QY	5940	AAGAAATAAGGGTATCAATACAGAGAGGAGTCAATATGCTCTGTTTCCAGATG	5999	Db	1042	rTyrSerSerAspLysGlyLeuIleSerArgIleTyrAsnGlnLeuLysGlnIleTyrLys	1082
Db	683	ysGluIleLysGlyIleGlnLeuGlyLysGluGluValLysLeuSerLeuPheAlaAspA	703	QY	7071	GAAGAAA---AACAAACCCCTCAAAATATGGGAAAAGATAGAGACACTTCTCAAA	7126
QY	6000	ACATGTTTGTATATTAGAAAACCCCTCTCTCAGGCCAAAACCTCTTAAAGCTGATAA	6059	Db	1062	sLysLysThrAsnAsnProIleLysLysTrpAlaLysAspMetAsnArgHisPheSerLys	1082
Db	703	spMetIleValTyrLeuGluAsnProIleValSerAlaGlnAsnLeuLysLeuLys	723	QY	7127	AGAAGACTTATCGACGCCAAACAAATATGAAAACCTCATCATCATCTGCTGCTTAG	7186
QY	6060	GCAACTTCAGAAAGTCTCAGGACACAAATCAATATGTCGAAAATTCACAGCTCTTAT	6119	Db	1082	sGluAspIleTyrAlaAlaLysLysHisMetLysCysSerSerSerLeuAlaIleAr	1102
Db	723	erAsnPheSerLysValSerGlyTyrLysIleAsnValGlnLysSerGlnAlaPheLeuT	743	QY	7187	AGAAATCGAAACAAACACACAGTGACATCCATCTCATGCTAGTAAATGGTGATCAC	7246
QY	6120	ACGCCAATATAGACAACAGAGACCAAAATCATGAGTGAACCTCTCATTCACAATTCCTA	6179	Db	1102	gGluMetGlnIleLysThrThrMetArgTyrHisLeuThrProValArgMetAlaIleI	1122
Db	743	yrThrAsnAsnArgGlnThrGluSerGlnIleMetGlyGluLeuProPheThrIleAlaL	763	QY	7247	TAAAGTCAGGAAACAAACAAATCTGGAGAGTATGGAGAAATAGGAACACTTTTCCA	7306
QY	6180	CAAGACAATAAATACCTAGGAATACAACTTACAAGGGACACGCTAGGAACTCTTCAAGG	6239	Db	1122	eLysLysSerGlyAsnAsnArgCysTyrArgGlyCysGlyGluIleGlyThrLeuLeuH	1142
Db	763	erLysArgIleLysTyrLeuGlyIleGlnLeuThrArgAspValLysAspLeuPheLysG	783	QY	7307	CTGTTGGTGGAAATGTAATTAGTTCAACCATTTGGAAGACAGCTGGAGATTCCTTAA	7366
QY	6240	AGAACTCAAAACCACTCATCAAGGAATAAGAGAGGACACAAACAAATGGAAAACATTC	6299	Db	1142	sCysTrpTrpAspCysLysLeuValGlnProLeuTrpLysSerValTrpArgPheLeuAr	1162
Db	783	luAsnTyrLysProLeuLeuLysGluIleLysGluAspThrAsnLysTrpLysAsnIleP	803	QY	7367	GGATCTAGAACCAACAAATATCATTTGACCCAGCAATCCCATTTACTAGTATATACCCAAA	7426
QY	6300	CATCCTCACAGATAGTAAGATCAT-----GAAATGGCATACTGCCCAAGTAATTT	6352	Db	1162	gAspLeuGluLeuGluIleProPheAspProAlaIleProLeuLeuGlyIleTyrProGl	1182
Db	803	roCysSerTrpValGlyArgIleAsnIleValLysMetAlaIleLeuProLysValIleT	823	QY	7427	GGATATAAATCTTCTTATTATAAGACACATGCACACATATGTTTATTCGACACTCAT	7486
QY	6353	ATAGATTGAGTGTACCCCTACCAAGCTACCATTTGACTTCTTCCACAGAATTTGAAAAA	6412	Db	1182	uAspTyrLysSerCysCysTyrLysAspThrCysThrArgMetPheIleAlaLeuPh	1202
Db	823	yrArgPheAsnAlaIleProIleLysLeuProMetThrPhePheThrGluLeuGluLysT	843	QY	7487	CACAAATAGCAAGACTTGGAAACCAACCAATGTCATCTAGTATAGACTGGGATAAGAA	7546
QY	6413	CAACTTAAATTTTCATATGGACCAAAAGAGCCACAGAGCCACAGCAATCTTAAAGC	6472	Db	1202	eThrIleAlaLysThrTrpAsnGlnProLysCysProThrMetIleAspTrpIleLysLy	1222
Db	843	hrThrLeuLysPheIleTrpAsnGln-LysArgAlaArgIleAlaLysSerIleLeuSer	862	QY	7547	AACATGCAATATACACCATGAATACTACTGACGCCATAAAG-CATGAGTTTCATGTC	7605
QY	6473	AAAAGAACAAAGCTGAGGTATCATCTACCTACCTTAAACTATATCTACTAAGCCTACA	6532	Db	1222	sMetTrpHisIleTyrThrMetGluTyrTyrAlaAlaIleLysAsnAspGluPheIleSe	1242
Db	863	GlnLysAsnLysAlaGlyGlyIleThrLeuProTyrPheLysLeuTyrTyrLysAlaThr	882	QY	7606	CTTTGACAGATATGATGAGCTGGAACCAACCATCATTTCTCAGCAAACTAACACAGAACA	7665
QY	6533	GTAACCAAACTGCTGCTACTGTGTACCAAAACAGATATATAGACCAATGGAACAGACA	6592	Db	1242	rPheValGlyThrTrpMetLysLeuGluThrIleIleLeuSerLysLeuSerGlnGluGl	1262
Db	883	ValThrLysThrAlaTrpTyrTrpTyrGlnAsnArgAspIleAspGlnTrpAsnArgThr	902	QY	7666	GAACCAACACACACATGTTCTCCTGTAAGTGGAGT 7705	
QY	6593	GAGACCTCAGAAATATAC-ACTGCAATCTACATCATCTGATCTTTTGACAAACCTGACAA	6651				

Db 1262 nLysThrLysHisArgIlePheSerLeuIleGlyGlyAsn 1275

RESULT 2

US-10-050-882-80

Sequence 80, Application US/10050882

Publication No. US2003010400A1

GENERAL INFORMATION:

APPLICANT: Ruben et al.

FILE OF INVENTION: 27 Human secreted proteins

FILE REFERENCE: P2038P1

CURRENT APPLICATION NUMBER: US/10/050,882

CURRENT FILING DATE: 2002-01-18

PRIOR APPLICATION NUMBER: 09/661,453

PRIOR FILING DATE: 2000-09-13

PRIOR APPLICATION NUMBER: PCT/US00/06783

PRIOR FILING DATE: 2000-03-16

PRIOR APPLICATION NUMBER: 60/125,055

PRIOR FILING DATE: 1999-03-18

NUMBER OF SEQ ID NOS: 156

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 80

LENGTH: 1010

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: SITE

LOCATION: (25)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (104)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (194)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (362)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (525)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (643)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (649)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (656)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (660)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (731)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (770)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (770)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (800)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (825)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (987)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

; NAME/KEY: SITE
; LOCATION: (996)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (1003)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-050-882-80

Alignment Scores:

Pred. No.: 0 Length: 1010
Score: 4081.50 Matches: 841
Percent Similarity: 76.80% Conservatives: 53
Best Local Similarity: 72.25% Mismatches: 111
Query Match: 7.80% Indels: 166
DB: 9 Gaps: 5

US-10-083-853B-2 (1-29921) x US-10-050-882-80 (1-1010)

QY 4244 ATGAAGGCAAAATAAAGATGTTCTTTGAACCAATGAGACAAAGACACAATGTACCAG 4303
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 MetLysAlaGluIleLysMetPheGluThrAsnGluAsnLysAspThrThrTyrGln 20
QY 4304 AATCTCTGGGCATATTAAACGAGTGTCTAGAGGAAATTTATAGCACTAGTGCCTAC 4363
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 21 AsnLeuTrpAsp***PheLysAlaValCysArgGlyLysPheIleAlaLeuAsnAlaHis 40
QY 4364 AAGAGAAAGCAGCAATATCTAAATAGACACCTTAAACATCACAATTAAGAACTAGAG 4423
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 41 LysArgLysGlnGluArgSerLysIleAspThrLeuThrSerGlnLeuLysGluLeuGlu 60
QY 4424 AAGAAGAGCAACAAATTCAAAAGCTAGCAGACAGCAAGAAATAACTAAGATCAGAGCA 4483
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 61 LysGlnGluGlnThrHisSerLysAlaSerArgGlnGluIleThrLysIleArgAla 80
QY 4484 GAACTGAAGGAGATAGACACACAAAAGCCCTTCAATAAATCAATGAATCAGAGCTG 4543
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 81 GluLeuLysGluIleGluThrGlnLysThrLeuGlnLys-IleAsnGluSerArgSerTr 100
QY 4544 GTTTTGTGAAAGATCAGCAAAAT---AGACCCTAGACAGACTAATAAACAAGAAAG 4599
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 100 PhePheGlu***IleAsnLysIleAspArgProLeuAlaArgLeuIleLysLysLys 120
QY 4600 AGAGAAGAAATCAAGAGATGCAATAAATAATGATAAAGGGGATATCACCCGATCCAC 4659
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 120 gLulLysAsnGlnIleAspAlaIleLysAsnAspLysGlyAspIleThrThrAspProth 140
QY 4660 AGAAATACAACTATTATCAGAGATATATATAACACCTCTATGCAATATAACTAGAAA 4719
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 140 rGluIleGlnThrThrIleArgGluTyrTrLysHisLeuTyrAlaAsnLysLeuGluAs 160
QY 4720 TCTAGAAGAAATGGATAAATTCCTGGACACATATGTAGCCTGTATGGACCTTGGGGGACA 4779
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 160 nLeuGluGluMetAspLysPheLeuAsp----- 169
QY 4780 GAACAAAGGGGTGAATGCAGAAATAAAGACAAAGACAAAGAGATGTTTGGAAAGTA 4839
Db 169 ----- 169
QY 4840 GGGGTTCAGGGGCAACTTGCCTCTAATGCACAAAGGCCCTGAGCTTACACCCCTCTG 4899
Db 169 ----- 169
QY 4900 TATTTATTAGGCAAAAGAGATAGCGAGAGGGTGAAGAGAGAGGTTCAGCTGTAGG 4959
Db 169 ----- 169
QY 4960 TCCAGAGTAGGCTGCAAGACTGCATTCCTCAACAATAGGCTCTAGATGTCCCAAGTACA 5019
Db 169 ----- 169
QY 5020 TAACCTCAAGGAGCAGTGCAGGAGGTGATGGCCCTCAGCAAACTTCTAGGCGAGCA 5079
Db 169 ----- 169

QY	5080	CAGAAGTAAGTTGGCCCAACATTCTGTATTTCAGATAAAACAGTTTGGCTGTTTGAACAGTA	5139
DB	169	-----	169
QY	5140	GCCTCCAGTGAATGCTGAGTTGGTTCATGATCCCTTTGGCCCTTTTGGCTCCCAAAACAC	5199
DB	170	-----	170
QY	5200	ATACACCCCTCTCAAGACTAAACACAGAGAAGTCAAAATCCCTGTAATATACAGTAACAAAG	5259
DB	170	rrYrThrLeuProArgLeuAsnGlnGluGluValGluSerLeuAsnArgProIleThrGI	190
QY	5260	TTCTAAATTAAGCAGTAATTTAGTCTACCAACCAAAAAAGTCCAGGACGACGG	5319
DB	190	ySerGluIle**AlaIleAsnSerLeuProThrLysLysSerProGlyProAspGI	210
QY	5320	ATTACAGCCAAATCTACACAGCTACAAAGAGAGCTGCTACTATTCTCTCGAACT	5379
DB	210	yPheThrAlaGluPheTyrglnArgTyrlsGluGluValProPheLeuLeuLysLe	230
QY	5380	ATTCCAAAAAATAGAA--AATGGGAATCCCTCCCTAACTCACTTTTACGAGCCACGATCA	5436
DB	230	uPheGlnSerIleGluLysGlu-GlyIleLeuProAsnSerPheTyrglnAlaSerIleI	250
QY	5437	TCCTGATACCAAACTAGCAGTACACAAACAAAAAGGAAATTTACAGGCCCATATCCC	5496
DB	250	leLeuIleProLysProGlyArgAspThrThrLysLysGluAsnPheArgProIleSerL	270
QY	5497	TGATGAACATTGATGGAATCCCTCAATAAAATCTGGCAACCAATCCAGCAGCAC	5556
DB	270	eumetAsnIleAspAlaLysIleLeuAsnLysIleLeuAlaAsnArgIleGlnGlnHisI	290
QY	5557	TCAAAAGCTTATCTACCATGATCAAGTTGGCGTCTATCCCTGGGATGCAAGCGTGTCA	5616
DB	290	leLysLysLeuIleHisAspGlnValGlyPheIleProGlyMetGlnGlytrpPhea	310
QY	5617	AAATATGCAAACTAATATGAGGCCATCATATAACAGAACCAATGACAAAAACACA	5676
DB	310	snIleArgLysSerIleAsnValIleGlnHisIleAsnArgThrLysAspLysAsnHisM	330
QY	5677	TGATTATCTCAATAGTCAGAAAGCGCTTGTCAAATTCACACAGCCCTTCATGCTAA	5736
DB	330	etIleIleSerIleAspAlaGluLysAlaPheAspLysIleGlnGlnProPheMetLeuL	350
QY	5737	AAATTCCTAGTAAGTATCGATGGAATGATCTCAAAATTAATAGAGCTATTATA	5796
DB	350	ysThrLeuAsnLysLeuGlyIleAspGlyThrTyrr**LysIleIleArgAlaIleTyra	370
QY	5797	C-AAACCCACAGCCATATCATATGAAATGGGCAAAAACTGGAAGCATTCCTCTTGAGAA	5855
DB	370	sPlysProThrAlaAsnIleLeuAsnGlyGlnLysLeuGluAlaPheProLeuLysT	390
QY	5856	CTGGCACACAGAGATGCCCTCTCTCACACACTCTTATTCAGATACTATTGGAAGTTC	5915
DB	390	hrGlyThrArgGlnGlyCysProLeuSerProLeuLeuPheAsnIleValLeuGluValI	410
QY	5916	TGGCCAGGCAATCGAGCAATAGAAGAAATAAAGGGTATTCAATAGAAAGAGAGGAAG	5975
DB	410	eulAaArgAlaIleArgGlnGluLysGluIleLysGlyIleGlnLeuGlyLysGluGluV	430
QY	5976	TCATATTTGCTCTGTTTCAGATGACATGTTTGTATATTATAGAAAAACCCATCTCTCAG	6035
DB	430	allLysLeuSerLeuPheAlaAspAspMetIleValTyrlLeuGluAsnProIleValSerA	450
QY	6036	GCCAAAAACTCTTAACTGATGAAGCACTTCAGCAAAAGTCTCAGGACACAAAATCAATG	6095
DB	450	laGlnAsnLeuLeuLysLeuIleSerAsnPheSerLysValSerGlyTyrlLysIleAsnV	470
QY	6096	TGCAAAATTCACAGCATTTATAGCCCAATATAGCAACACACAGAGCCAAATCATGA	6155
DB	470	alGlnLysSerGlnAlaPheLeuTyrrAsnAsnArgGlnThrGluSerGlnIleMets	490

```

Db      849  uThrProValArgMetAlaIleLeuLysSerGlyAsnAsnArgCysTrpArgGlyCy 869
QY      7283  TGGAGAAATAGAACACATTTTCCACTGTTGGTGGGAATGTAATAGTTACCACTATGG 7342
Db      869  sGlyGluIleGlyThrLeuLeuHisCysTrpArgCysLysLeuValGlnProLeuTr 889
QY      7343  GAAGACAGTGTGGAGATTCCTTAAGGATCTAGAACACAGAAATATCATTTGACCCAGCAAT 7402
Db      889  pLysSerValTrpArgPheLeuArgAspLeuGluLeuGluIleProPheAspProAlaIle 909
QY      7403  CCATTACTGAGTATATACCCAAAGGAATATATAATCATCTTATTATAAGACACATGCAC 7462
Db      909  eProLeuLeuGlyIleTyProLysAspTyLysSerCysCysTyLysAspThrCysTh 929
QY      7463  ACATATGTTTATTGAGCAGCTCATCACATACAGAACTTGGAAACCAACCCAAATGTC 7522
Db      929  rArgMetPheIleAlaLeuPheThrIleAlaLysThrTrpAsnGlnProLysCysPr 949
QY      7523  ATCAGTGTAGACTGGATAAGAAACATATGACATATACACCATGAATACTATGCAGC 7582
Db      949  oThrMetIleAspTrpIleLysLysMetTrpHisIleTyThrMetGluTyTrpAlaAl 969
QY      7583  CATAAAG - GATGAGTTCATCTCTTCGACAGATATGATGAGAGCTGGAACCATCAT 7641
Db      969  aIleLysAsnAspGluPheMetSerPheValGlyThrTrpMetLysLeuGlu***IleI 989
QY      7642  TCTCAGCAAACTAACACAGAAACAAACCAACCAACCATGTTCTCACTTGTAAAGTGG 7701
Db      989  eLeuSerLysLeuSerGln**GlnLysThrLysHisArg***PheSerLeuIleGlyGl 1009
QY      7702  GAGT 7705
Db      1009  yAsn 1010

RESULT 3
US-10-000-256A-153
; Sequence 153, Application US/10000256A
; Publication No. US2003003983A1
; GENERAL INFORMATION:
; APPLICANT: Sun, Yongming
; APPLICANT: Recipon, Herve
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Prostate Specific Genes and
; FILE REFERENCE: DEX-0259
; CURRENT APPLICATION NUMBER: US/10/000,256A
; CURRENT FILING DATE: 2001-11-01
; PRIOR APPLICATION NUMBER: 60/244,782
; PRIOR FILING DATE: 2000-11-01
; NUMBER OF SEQ ID NOS: 240
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 153
; LENGTH: 956
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-000-256A-153

Alignment Scores:
Pred. No.: 1,95e-297 Length: 956
Score: 3770.50 Matches: 783
Percent Similarity: 72.91% Conservative: 65
Best Local Similarity: 67.33% Mismatches: 104
Query Match: 7.21% Indels: 218
DB: 9 Gaps: 7

US-10-083-853B-2 (1-29921) x US-10-000-256A-153 (1-956)
QY      4244  ATGAGGCAAAATTAAGATGTTCTTTGAAACCAATGAGAACAAAGACACAATGTACCAG 4303
Db      1  MetLysAlaGluIleLysValPhePheGluThrAsnGluAsnLysAspThrThyTrpGln 20
QY      4304  AATCTCTGGGGCATATTAAAGACAGTGTGTAGAGGGAAATTTATAGCACTAGATGCCTAC 4363
```

```

Db      21  AsnLeuTrpAspThrPheLysAlaValCysArgGlyLysPheIleAlaLeuAsnAlaHis 40
QY      4364  AAGAGAAAGCAGGAATATCTAAATATAGACACCTTTAAATCATCACTTAAAGAACTAGAG 4423
Db      41  LysArgLysGlnGluArgSerLysIleAspThrLeuThrSerGlnLeuLysGluLeuGlu 60
QY      4424  AAGAAAGAGCAACAAATTTCAAAGCTAGCAGAGACAGAAATAACTAAGATCATCAGACA 4483
Db      61  LysGlnGluGlnThrHisSerLysAlaSerArgArgGlnGlnIleThrLysIleArgAla 80
QY      4484  GAACCTGAGGAGATAGACACACAAAGCCCTTCAATTAATCAATGAATCATTCAGAGCTG 4543
Db      81  GluLeuLysGluIleGlnThrGlnLysThrLeuGlnLys - IleAsnGlnSerArgSerTr 100
QY      4544  GTTTTTTGAAGAGATCAGCAAAATAGACCA ----CTAGACAGACTTAATAAAGAAAGAG 4599
Db      100  pPheGluArgIleAsnLysIleAspSerLeuAlaArgLeuIleLysLysLysAr 120
QY      4600  AGAGAGAAATCAAGAGATGCAATATAAATAATGATAAGGGGATATACACCCGATCCAC 4659
Db      120  gLulLysAsnGlnIleAspThrIleLysAsnAspLysGlyAspIleThrThrAspProTh 140
QY      4660  AGAAATACAAACTATTATCAGAGAAATATATAACACCTCTATGCAATATAACTAGAAAA 4719
Db      140  rGluIleGlnThrThrIleArgGluTyTrpLysHisLeuTyAlaAsnLysLeuGluAs 160
QY      4720  TCTAGAAAGATGGATAATTCCTCGACACATATGTAGCCTGTATGGACCTTGGGGGACA 4779
Db      160  nLeuGluGluMetAspLysPheLeuAsp ----- 169
QY      4780  GAACAAAGGGGGTGAATGCAGAAATAAAGACAAAGAGAGATATGTTTGAAGTA 4839
Db      169  ----- 169
QY      4840  GGGGTTCAGGGGGCAACTTGCCTCTTAATGGACAAGGGCCCTGAGCTTTACACCACCTCTG 4899
Db      169  ----- 169
QY      4900  TATTTTATAGCAAAAGAGATAGCGAGAGGGTGAGTTGGAGAAGAGGTTCAGCTCTTAGG 4959
Db      169  ----- 169
QY      4960  TCCAGAGTAGCCCTGCAAGACTGCATTCCTCAAAACAATAGGCTCTAGATGTCCCACTAGA 5019
Db      169  ----- 169
QY      5020  TAACCTCAAGGAGCCAGTGCAGGGAGTGATGGCCCTCAGCAAAACCTTCTAGGGCAGCGA 5079
Db      169  ----- 169
QY      5080  CAGAAAGTAAGTTTGCCCCACATTTCTATTACGATAAACAGTTTGTCTGTTGATCAAGTA 5139
Db      169  ----- 169
QY      5140  GCCTCCAGTGAATGCTGAGTTGGTGCATGATGCCTTTTGGCCCTTTTGGCTCCCAAAACAC 5199
Db      170  -----Th 170
QY      5200  ATACACCCTCTCAAGACTAAACACAGGAAGAGTCAATCCCTGATATATACCACTAACAG 5259
Db      170  rTyThrLeuProArgLeuAsnGlnGluValGluSerLeuAsnArgProIleThrGl 190
QY      5260  TTCTAAATTCAGCAGTAATTTGATAGCTTACCACCAAAAAAGTCCAGGACCAAGCGG 5319
Db      190  yAlaGluIleValAlaIleIleAsnSerLeuProThrLysLysSerProGlyProAspGl 210
QY      5320  ATTCACAGCCAAATTTCTACCAAGAGGTACAAAGAGAGAGTGTGCTACTATTCTTCTGAAACT 5379
Db      210  yPheThrAlaGluPheTyTrpGlnSerTrpAlaGlu ----- 221
QY      5380  ATTCCAAAAAATAGAAAAATGGGAATCCTCCCTTAACCTATCTTTACGAGGCCACATCATCC 5439
```

```
Db 221 ----- 221
Qy 5440 TGATACAAAACCTAGCAGTCACACAAACAAAGAGAAATTTTCAGGCCATATCCCTGA 5499
Db 222 ---Thr-GlnPro-----LysLysGluasnPheargProIleSerLeuM 235
Qy 5500 TGAACATTGATGTGAAATCCTCAATAAATCTGCGAAACCAATCCAGCAGCACATCA 5559
Db 235 etAsnIleAspAlaLysIleLeuAsnLysIleLeuAlaLysArgIleGlnHisIleL 255
Qy 5560 AAAAGCTTATCTACCATGATCAAGTTGGCGTCATCCCTGGGATCGAAGCTGTTCACAAA 5619
Db 255 yLysLeuIleHisIleAspGlnValGlyPheIleProGlyMetGlnGlyTrpPheAsnI 275
Qy 5620 TATCAAAATCAATAATGTAGGCCATCACATAAAGACAAACCAATGACAAAACACACATGA 5679
Db 275 leArgLysSerIleAsnValThrGlnHisIleAsnArgAlaLysAspLysAsnHisMetI 295
Qy 5680 TTATCTCAATAGATGCAGAAAGCCCTTGTCAAAATTCACAGCCCTTCATGCTCAAAA 5739
Db 295 leIleSerIleAspAlaGluLysAlaPheAspLysIleGlnGlnProPheMetLeuLysT 315
Qy 5740 TTCTCAGTAACTAGGTATCGATGGAATGTATCTCAAAATTAAGAGCTATTTATAC-A 5798
Db 315 hrLeuAsnLysLeuGlyIleAspGlyThrTrpPheLysIleIleArgAlaIleTrpAspA 335
Qy 5799 AACCCACAGCAATATATCATCTGAATGGCAAAACCTGGAAGCAATTCCTTTGAGAAGCTG 5858
Db 335 snProThrAlaAsnIleIleLeuAsnGlyGlnLysLeuGluAlaPheProLeuLysThrG 355
Qy 5859 GCACAAGCAAGGATGCCCTCTCACACTCTTATTCAGATACTACTTGAAGTCTTGG 5918
Db 355 lyThrArgGlnGlyProLeuSerProLeuLeuPheAsnIleValLeuGluValLeuA 375
Qy 5919 CCAGGCAATCAGCAATAGAAAGCAATTAAGGGTATTCAAAATAGAAAGAGCAAGTCA 5978
Db 375 laArgAlaIleArgGlnGluLysGluIleLysGlyIleGlnLeuGlyGluValL 395
Qy 5979 TATTGTCTCTGTTGCAGATGACATGTTGTGTATATTAGAAAACCCCATCTCTCAGGCC 6038
Db 395 ysLeuSerLeuPheAlaAspAsnMetIleValTyLeuGluasnProIleValSerAlaG 415
Qy 6039 AAAAATCTTAACTGATAGCAACTTCACCAAGTCTCAGGACACAAAATCAATGTGC 6098
Db 415 InAsnLeuLeuLysIleSerAsnPheSerLysValSerGlyTyLysIleAsnValG 435
Qy 6099 AAAAATCACAAGCATCTTATACGCCAATAATACACAAACAGACAGCAATCATGAGTG 6158
Db 435 InLysSerGlnAlaPheLeuTyThrAsnAsnArgGlnThrGluSerGlnIleMetSerG 455
Qy 6159 AACTCTCATTCACAAATTCACAAAGAGATAAATAATACCTAGGAATACAACTTACAAGGG 6218
Db 455 InLeuProPheThrIleAlaSerLysArgIleLysTyLeuGlyIleGlnLeuThrArgA 475
Qy 6219 ACAGGTAGAACTCTCAAGGAGAACTACAAACACTGATCAAGGAATAAGAGAGACA 6278
Db 475 spValLysAspLeuPheLysGluAsnTyLysProLeuLeuLysGluIleLysGluAspT 495
Qy 6279 CAACAATGGAACAAATCCATGCTCACAGATAGTAAGATCAT-----GAAATG 6331
Db 495 hrAsnLysTyTrpLysAsnIleProCysSerGlyGluGlyArgIleAsnIleValLysMetA 515
Qy 6332 CCATACCTGCCAAAGTAAATATATAGATTGCTAGTGTACCCCATCAAGCTACCATGACTT 6391
Db 515 lalleLeuProLys----- 519
Qy 6392 TCITTCACAGAAATGGAAAAACAACTTTAAATTTTCATATGGAACCAAAAAAGAGCCAC 6451
Db 520 -----GluLeuGluLysThrThrLeuLysPheIleTrpAsnGln-LysArgAlaHis 536
Qy 6452 AGACCCAGACAATCTTAAGCAAAAACAAACAACTCGAGGTATCATGCTACCTGACTTA 6511
Db 537 IleAlaLysSerIleLeuAsnGlnLysAsnLysAlaGlyGlyIleThrLeuProAspPhe 556
```

```
Qy 6512 AAATATACATATAGGCTACAGTAACCAAACTGCATGCTACTGCTACCAAAACAGATAT 6571
Db 557 LysLeuTyTrpLysAlaThrValThrLysThrAlaTrpTyTrpTyGlnAsnArgAsp 576
Qy 6572 ATAGACCAATGGAACAGACAGACACCTCAGAAATTCACACT-GCAATCTACATCCATCTG 6630
Db 577 IleAspGlnTrpAsnArgThrGluProSerGluIleThrGlnHisIleTrpSerTyLeu 596
Qy 6631 ATCTTTTCCAAACCTGACAAAACAAACCAATGGAAGGATTCCTTATTAAATAATG 6690
Db 597 IlePheAspLysProGluLysAsnLysGlnTrpGlyLysAspSerLeuPheAsnLysTrp 616
Qy 6691 TGTGTGAAAAAATCGCTAGCATATGAGAAAGCTGAAACTGGATCCCTTCTTACACCT 6750
Db 617 CysTrpGluAsnTrpLeuAlaIleCysArgLysLeuLysLeuAspProPheLeuThrPro 636
Qy 6751 TATCAAAAGTTAACTCAAGATGAATTAAGACATTAATATATAAGACATTAACACATAAAA 6810
Db 637 TyThrLysMetAsnSerArgTrpIleLysAspLeuAsnValArgProLysThrIleLys 656
Qy 6811 ACCCA-GAAGAAAACCTAGGCAATACCATTCAGCATATGGCATGGGCAAGACTTCATG 6869
Db 657 ThrLeuGluGluAsnLeuGlyIleThrIleGlnAspIleGlyMetGlyLysAspPheMet 676
Qy 6870 ACTAAAAACCAAAAGCAATGGCAACAAAGCCCAAAATAGACAAAGTGGATCTGATTAAA 6929
Db 677 SerLysThrProLysAlaMetAlaThrLysAspLysIleAspLysTrpAspLeuValLys 696
Qy 6930 CTATAGACCTCTCCACAGCAAAAACAAACTGCTCATCAGAGTGAACAGCAACCTACAGA 6989
Db 697 LeuLysSerPheCysThrAla-LysGluThrIleArgValAsnArgGlnProThrLy 716
Qy 6990 ATGGGAGAAAATTTTCAATCTATCGATCTGACAAAGGCTAATATCAGAGATCTACGA 7049
Db 716 stpGluLysIlePheAlaThrTySerSerAspLysGlyLeuIleSerArgIleTyAs 736
Qy 7050 AGAACTTAAACAAATTTACAAAGAAAAA-----AACAAACCCCTCAAAATATGGCAAGGA 7105
Db 736 nGluLeuLysGlnIleTyLysLysLysThrAsnAsnProIleLysLysTrpAlaLysAs 756
Qy 7106 TATCAGCAGACACTTCTCAAAAGAGACATTTATGACGCCAACCAACAAATATGAAAAAAC 7165
Db 756 pMetAsnArgHisPheSerLysGluAspIleTyAlaAlaLysLysHisMetLysLysCy 776
Qy 7166 CTATCATCATCTGCTGCTAGAGAAATGCAAAACAAACCAACACAGTGCATACCATCTCAT 7225
Db 776 sSerSerSerLeuAlaIleArgGluMetGlnIleLysThrThrMetArgTyHisLeuTh 796
Qy 7226 GCTAGTTAGATGCTGCTACTTAAAGAGTCAGGAAACAAACAAATGCTGGAGAGATGTGG 7285
Db 796 rProValArgMetAlaIleIleLysLysSerGlyAsnAsnArgCysTrpArgGlyCysGl 816
Qy 7286 AGAAATAGGACACTTTCCACTGTTGGTGGGAATGAATTAATAGTTCAACCATCTGTGAA 7345
Db 816 yGluThrGlyThrLeuLeuHisCysTrpTrpAspCysLysLeuAlaGlnProLeuTrpLy 836
Qy 7346 GACAGTCTGGAGATTCCTTAAGGATCTAGAACCAAGAAATATCATTTGACCCAGCAATCCC 7405
Db 836 sSerValTrpArgPheLeuArgAspLeuGluLeuGluIleProPheAspProAlaIlePr 856
Qy 7406 ATTACTGAGTATATACCAAGGAATATAAATCTATTCTATTATAAGACACATCGACACA 7465
Db 856 oLeuLeuGlyIleTyTrpLysAspTyTrpLysSerCysCysTyTrpLysAspThrCysThr 876
Qy 7466 TATGTTTATGACACTGATCACAATAGCAAGACTTGGAAACCAACCAATGTCATC 7525
Db 876 gMetPheIleAlaAlaLeuPheThrIleAlaLysThrTrpAsnGlnProLysCysProTh 896
Qy 7526 AGTCATAGACTGGATAAAGAAACATGGCACATATACACATGAATGAATACTATCAGCCAT 7585
Db 896 rIleIleAspTrpIleLysLysMetTrpHisIleTyThrMetGluTyTrpAlaAlaI 916
```



```
QY 5581 AAGTGGCGCATCCCTGGGATGCAAGCGCTGGTTCAAAATATATGCAAAATCAATAATGTAG 5640
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
352 InValGlyPheIleProGlyMetGlnGlyTrpPheAsnIleArgLysSerIleAsnValI 372
QY 5641 GCCATCATAAACAGAACCAATGACAAAACACACATGATTCTCAATAGATGACAGAA 5700
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
372 leGlnHisIleAsnArgAlaLysAspLysAsnHisMetIleIleSerIleAspAlaGluL 392
QY 5701 AGGCCTTTGTCAAAATTCACAGCCCTTCATGCTTAAATTCCTCACTAAACTAGCTATCG 5760
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
392 ysaIaPheAspLysIleGlnGlnProPheMetLeuLysThrLeuAsnLysLeuGlyIleA 412
QY 5761 ATGGAATGCTATCTCAAAATATAAGAGCTATTATATAC-AAACCCACAGCCCAATATCATAC 5819
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
412 spGlyThrTyPheLysIleIleArgAlaIleTyAspLysProThrAlaAsnIleIleL 432
QY 5820 TGAATGGGCAAAACTGGAGCATTCCTTTGAGAACTGGCAAGCAAGAGATGCGCTC 5879
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
432 euAsnGlyGlnLysLeuGluAlaPheProLeuLysThrGlyThrArgGlnGlyCysProL 452
QY 5880 TCTCACCACCTCTATTCAAGATACTATTGGAAGTCTCTGGCCAGGCAATCAGGCAATAGA 5939
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
452 euserProLeuLeuPheAsnIleValLeuGluValLeuAlaArgAlaIleArgGlnGluL 472
QY 5940 AAGAAATPAAAGGTATTCAATAGAAAAGAGAGGAAGTCATATTCTCTCTTTGCGAGTG 5999
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
472 ysgLulLysGlyIleGlnLeuGlyLysGluGluValLysLeuSerLeuPheAlaAspA 492
QY 6000 ACATGTTGTATATTAGAAAACCCATCTCTCAGGCCAAAACCTCTTAAGCTGATAA 6059
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
492 spMetIleValTyLeuGluAsnProIleValSerAlaGlnAsnLeuLeuLysLeuIleS 512
QY 6060 GCAACTTCAGCAAGTCTCAGGACACAAATCAATGTGCAAAATCACAGCATCTCTAT 6119
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
512 erAsnPheSerLysValSerGlyTyTrLysIleAsnValGlnLysSerGlnAlaPheLeu 532
QY 6120 ACGCCAATAATAGACAACAGAGCCAAATCATAGTGAACCTCTCATTCACAAATGCTA 6179
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
532 yThrAsnAsnArgGlnThrGluSerGlnIleMetSerGluLeuProPheThrIleAlaS 552
QY 6180 CAAGAGAAATAAATACCTAGGATACAACTTACAGGGACACGTAGGAACCTCTCAAGG 6239
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
552 erLysArgIleLysTyLeuGlyIleGlnLeuThrArgAspValLysAspLeuPheLysG 572
QY 6240 AGAACTCAACACCTCATCAAGSAAATAGAGAGGACACAAACAATGCAAAACATTC 6299
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
572 luAsnTyLysProLeuLeuLysGluIleLysGluAspThrAsnLysTrpLysAsnIleP 592
QY 6300 CATGCTCACAGATAGTAAGAAATCAT-----GAAATGCCATACTGCCCAAGTAAT 6352
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
592 rocYsSerTrpValGlyArgIleAsnIleValLysMetAlaIleLeuProLysValIleT 612
QY 6353 ATGATTCAGTGTACCCCATCAAGTACCATGCTTCTTCCACAGAAATGGAAGAAA 6412
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
612 yrArgPheAsnAlaIleProIleLysLeuProMetThrPhePheThrGluLeuGluLysT 632
QY 6413 CAACTTTAAATTCATATGAACCAACAAAAGACCCACAGAGCCAGACAACTTAAGC 6472
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
632 hrThrLeuLysPheIleTrpAsnGln-LysArgAlaArgIleAlaLysSerIleLeuSer 651
QY 6473 AAAAGAACAAAGCTGGAGGTATCATGCTACCTGACTTAAAACTACTACTAAAGGCTACA 6532
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
652 GlnLysAsnLysAlaGlyGlyIleThrLeuProAspPheLysLeuTyTrLysAlaThr 671
QY 6533 GTAACCAAACTGAGGTACTGCTACCAACAAACAGATATATAGACCAATGGACAGACA 6592
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
672 ValThrLysThrAlaTrpTyTrTyTrGlnAsnArgAspIleAspGlnTrpAsnArgThr 691
QY 6593 GAGACCTCAGAAATTAC-ACITGCAATCTACATCCATCTGATCTTTGACAACTGACAAA 6651
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
692 GluProSerGluIleMetProHisIleTyAsnTyLeuIlePheAspLysProGluLys 711
```

```
QY 6652 AACAGCAATGGAAGGATTCCTTATTAATAATGTTGTTGGAAAAAAGTGGCTAGCC 6711
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
712 AsnLysGlnTrpGlyLysAspSerLeuPheAsnLysTrpCysTrpGluAsnTrpLeuAla 731
QY 6712 ATATGAGAAAGCTGAACCTGGATCCCTTCTTACACCTTATACAAAAGTTAACTCAGA 6771
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
732 IleCysArgLysLeuLysLeuAspProPheLeuThrProTyThrLysIleAsnSerArg 751
QY 6772 TGAATTAAGACACTTAATATAACACATAAAACCATAAAAACCCA-GAAGAAAAACCTAGGC 6830
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
752 TrpIleLysAspLeuAsnValArgProLysThrIleLysThrLeuGluGluAsnLeuGly 771
QY 6831 AATACCATTCAGGATATGGACATGGCCAAAGACTTTCATGACTAAAACACCAAGCAATG 6890
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
772 IleThrIleGlnAspIleGlyValGlyLysAspPheMetSerLysThrProLysAlaMet 791
QY 6891 GCAACAAAACCCAAATAGACAAGTGGGATCTCTATTAAACTATAGAGCTTCTGCACAGCA 6950
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
792 AlaThrLysAlaLysIleAspLysTrpAspLeuIleLysLeuLysSerPheCysThrAla 811
QY 6951 AAAAAAACTGTCTCAGAGTGAACCAAGCAACCTACAGAATGGAGAAAAATTTTGCAT 7010
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
812 -LysGluThrThrIleArgValAsnArgGlnProThrThrTrpGluLysIlePheAlaTh 831
QY 7011 CTATGATCTGCAAAAGGCTTAATATCCAGAGATCTACGAAGAAGTCTTAAACAAAATTTACAA 7070
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
831 rTySerSerAspLysGlyLeuIleSerArgIleTyAsnGluLeuLysGlnIleTyL 851
QY 7071 GAAAAA-----AACACCCCTCAAAATATGGGCAAGGATATAGGACAGACTTCTCAAA 7126
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
851 sLysLysThrAsnAsnProIleLysLysTrpAlaLysAspMetAsnArgHisPheSerL 871
QY 7127 AGAAGACATTATGCGCCCAACACATATGAAAAAACCTCATCATCTGTCGTTAG 7186
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
871 sGluAspIleTyAlaAlaLysLysHisMetLysLysCysSerSerLeuAlaIleAr 891
QY 7187 AGAAATGCAAAACAAACACAGTGCATACCATCTCATCTAGTTAGTATGTTGATCAC 7246
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
891 gGluMetGlnIleLysThrMetArgTyHisLeuThrProValArgMetAlaIleIl 911
QY 7247 TAAAAAGTCAGGAACCAAC 7265
Db      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
911 eLysLysSerGlyAsnAsn 917
RESULT 5
US-10-082-830-235
; Sequence 235, Application US/10082830
; Publication No. US20030077604A1
; GENERAL INFORMATION:
; APPLICANT: Sun, Yongming
; APPLICANT: Recipon, Herve
; APPLICANT: Salceda, Susana
; APPLICANT: Liu, Chenghua
; APPLICANT: Turner, Leah
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific
; FILE REFERENCE: DEX-0249
; CURRENT APPLICATION NUMBER: US/10/082, 830
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/243, 802
; PRIOR FILING DATE: 2000-10-27
; NUMBER OF SEQ ID NOS: 282
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 235
; LENGTH: 940
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-082-830-235
Alignment Scores: 2.58e-278 Length: 940
Pred. No.: 3535.50 Matches: 747
Score: 74.43%
Percent Similarity: 68
```

Best Local Similarity: 68.22% Mismatches: 119
Query Match: 6.76% Indels: 167
DB: 9 Gaps: 7

US-10-083-853B-2 (1-29921) x US-10-082-830-235 (1-940)

```
QY 4016 GAATATACACTCTCTCAGCATCATACACTTAAATTTAAATTTGACCATGATATTTTA 4075
||||| ||||||| || ||||||| ||||||| ||||||| ||||||| |||||||
Db 1 GluTyrThrSerPheSerAlaLeuHisAsnThrTyrSerLysIleAspHisIleValGly 20
QY 4076 ACTAAACACTCTCTCAGCAAGTCAAAAGACAGAAATCCCTAACAAACAGTCTCTCAGAC 4135
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 21 SerLysAlaLeuSerLysCysLysArgGluMetIleThrAsnCysLeuSerAsp 40
QY 4136 TACAGTGAATCTATTAGAACTCAGAAATTAAGAACTCACCCTCAAAATCACAACTACA 4195
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 41 HisSerAlaIleLysLeuGluLeuArgIleLysLysLeuThrGlnAsnCysSerThrThr 60
QY 4196 TGGAACTGAACAACCTGCTCTGAATGACTACTGGTAAATTAACAAATGAAGCAAAA 4255
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 61 TrpLysLeuAsnAsnLeuLeuLeuAsnAspTyrCysValHisAsnLysMetLysAlaGlu 80
QY 4256 ATAAAGATGCTCTTGAACCAATGAGAACAGACACAATGTACCAGAACTCTCGGGC 4315
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 81 IleLysMetPhePheGluThrAsnGluAsnLysAspThrThrTyrGlnAsnLeuTrpAsp 100
QY 4316 ATATTAAAGCAGTGTAGAGGAAATTTATAGCACTAGATGCCCTACAAGAGAAAGCAG 4375
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 101 ThrPheLysAlaValCysArgLysAsnPheIleAlaLeuAsnValHisLysArgLysGln 120
QY 4376 GAAATATCTAAATAGACACCTTAACATCAATTAAGAACTAGAGAAAGAGAGCAA 4435
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 121 GluArgSerLysIleAspThrLeuIleSerGlnLeuLysGluLeuGluLysGlnGluGln 140
QY 4436 ACAAAATCAAAAGCTAGCAGAGACAAAGAAATACTAGATCAGACAGAACTGAGAGAG 4495
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 141 ThrHisSerLysAlaSerArgArgGlnGluIleThrLysIleArgAlaGluValLysGlu 160
QY 4496 ATAGAGACAAAAAGCCCTCAAAATAATCAATCAATCCAGGAGCTGTTTGAAGA 4555
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 161 IleGluThrGlnLysThr-PheLysArgIleAsnGluSerArgAsnTrpPhePheGluAr 180
QY 4556 GATCAGCAAAAT---AGACCACTAGACAGACTAATAAGAGAAAGAGAGAGAAATCA 4611
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 180 GileSerLysIleAspArgProLeuAlaArgLeuIleLysLysLysArgGluLysAsnG 200
QY 4612 AAGAGATCAATAAAAA---AATGATAAAGGGGATATCACCCGATCCACAGAAATACA 4668
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 200 nileAspAlaIleAsnThrHisAspLysGlyAspIleThrAspProThrGluIleG 220
QY 4669 AACTATTATCAGAGAAATATTATAACACCTCTATGCAAAATAACTAGAAAAATCTAGAAGA 4728
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 220 nThrThrIleArgGluTyrTyrLysHisPheTyrAlaAsnLysLeuGluAsnLeuGluG 240
QY 4729 AATGATAAATCTCGGACACATATGTAGCCTGTATGACCTTGGGGGACAGACAAAG 4788
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 240 uMetAspLysPheLeuAsp----- 246
QY 4789 GGGGTGAATGCAGAAATAAAGACAAACAGAGATATGTTGGAAGTAGGGGTGAGG 4848
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 246 ----- 246
QY 4849 GGGCAACTTGCCCTTAATGGACAGGGCCCTGAGCTTTACACCACCCCTCTGTATTATTA 4908
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 246 ----- 246
QY 4909 GCAAAAGAGATAGCGAGGGGTGAGTTGGAAGAGAGGTCAGCTGTAGTCCAGAGTA 4968
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 246 ----- 246
QY 4969 GGCCTGCAAGACTGCATTCCTCAAAACANTAGCCTCTAGATGTCCAGTAGATAACCTCAA 5028
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 246 ----- 246
```

```
QY 5029 GGAGCCAGTGCCAGGAGTGTATGGCCCTCAGCAAACTTTCTAGGCGAGGCACAGAGTAA 5088
246 ----- 246
QY 5089 GTTTGCCCCACATCTCTGTATTACAGATAAACAGTTGTGCTTGTGATCAAGTAGCTCCAGT 5148
246 ----- 246
QY 5149 GGAATGCTGAGTTGGTCATGATCCCTTTGGCCTTTTGGCTCCCAAAACACATACACCT 5208
247 ----- 247
QY 5209 CTCAGACTAAACAGGAGAGTCAAAATCCCTGAATATACCAAGTAAACAGTCTTAAAT 5268
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 250 uProArgLeuAsnGlnGluAlaGluSerLeuAsnArgProIleThrAspSerGluI 270
QY 5269 TGAAGCAGTAATGTATGACCTACCAACCAAAAGTCCAGGACAGGATTCACAGC 5328
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 270 eAlaAlaIleIleAsnSerLeuProThrLysLysSerProGlyProAspGlyPheThrPr 290
QY 5329 CAATTTCTACAGAGGTACAAAGAGAGCTGTACTATTCTCTGAAACTATTCCAAAA 5388
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 290 oLysPheTyrGlnArgTyrLysGluLeuValProPheLeuLeuLysLeuPheGlnSe 310
QY 5389 AATA---GAAATGGGAATCTCCCTAACTATTTACGAGCGCCAGCATCTCTGTATAC 5445
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 310 rIleThrLysGlu-GlyIleLeuProAsnSerPheTyrGluAlaAsnIleIleLeuIleL 330
QY 5446 CAAACTAGCAGTGACACACAAAAAGAGAAATTTCCAGGCCCTATCCCTGATGACAA 5505
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 330 euLysProGlyArgAspThrThrLysLysArgGluPheArgProIleSerMetMetIleI 350
QY 5506 TTGATGCAAAATCCTCAATAAAATACTGCAAAACCAATCCAGCAGCACATCAAAAAGC 5565
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 350 leAspAlaLysIleLeuSerLysIleLeuAlaAsnGlnIleGlnHisLeuIleLysL 370
QY 5566 TTATCTACCATGATCAAGTTGGCGTCACTCCCTGGGATGCAAGGCTGGTTCAAAATATCA 5625
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 370 euIleHisAspGlnValGlyPheIleProGlyMetLysGlyTrpPheAsnIleArgL 390
QY 5626 AATCAATAAATGTAGGCCATCACATAAACACAGAACCAATGACAAACACACATGATATCT 5685
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 390 ySerIleLysValIleHisIleAsnArgThrLysAspLysAsnHisMetIleIleI 410
QY 5686 CAATAGATGCAAGAAAGCCCTTTGTCAAAATTCACAGCCCTTCATGCTAAATATCTCA 5745
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 410 erIleHisAlaGluLysAlaPheAspLysIleGlnGlnProPheMetLeuLysThrValA 430
QY 5746 GTAACTAGGTATCGATGGAATGTATCTCAAAATAAAGAGCTATTATTATAC-AAACCCA 5804
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 430 snLysLeuValIleAspGlyThrTyrLeuLysIleIleArgAlaIleTyrAspLysProt 450
QY 5805 CAGCCAATATCATACTGNAATGGGCAAAACCTGGAGCATTCCTCTTGAGACTGGCACAA 5864
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 450 hrAlaAsnIleIleLeuAsnGlyGlnLysLeuGluAlaPheProLeuArgThrGlyLea 470
QY 5865 GACAAGGTGCGCTCTCTCACACTCTCTATTCAAGATACTATTGGAAGTTCTGGCCAGGG 5924
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 470 rGlnGlyCysProLeuSerProLeuPheAsnIleValleuGluValleuAlaArgA 490
QY 5925 CAATCAGCAATAGAGAAATAAAGGCTATTCAAAATAGAAAGAGAGAGGAGTCAATTTGT 5984
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 490 laIleArgGlnGluLysGluIleLysGlyIleGlnLeuGlyLysGluLysValLysLeuS 510
QY 5985 CTCCTGTTGCAGATGACATGTTGTATATTTAGAAAACCCCATCGTCTCAGGCCCAAAAC 6044
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 510 erLeuPheAlaAspMetIleTyrLeuGluAsnProIleValSerAlaGlnAsnL 530
QY 6045 TCCTTAAGCTCATAGCAACTTCAGCAAAAGTCTCAGGACACAAATCAATCTGCAAAAAT 6104
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 530 euLeuLysLeuMetSerSerPheSerLysValSerGlyTyrLysIleAsnValGlnLysS 550
```

```
QY 6105 CACAAGCATTTCTTATAGCCCAATAATAGACAACAGAGAGCCAAATCATGAGTGAATCT 6164
Db 550 erGlnAlaPheLeuThrThrAsnAsnArgGlnThrGluSerGln---MetSerGluLeuP 569
QY 6165 CATTCAATTTGCTACAAAGAGATAAATAATACCTAGGAATACAACTTACAGAGGACAGCT 6224
Db 569 roPheAlaIleAlaSerLysArgIleLysThrLeuGlyIleGlnLeuThrArgAspValL 589
QY 6225 AGGAATCTTCAAGAGAGAACTACAAACCACTGATCAAGGAATAAGAGAGGACACAAACA 6284
Db 589 ysAspLeuPheLysGluAsnThrLysProLeuLeuAsnLysLysLysGluAspThrAsnL 609
QY 6285 AATGGAATAACATTCCTGCTACAGATAGTAAGAAATCAT-----GAAATGCCATAC 6337
Db 609 ystrLysAsnIleProCysSerTrpIleGlyArgIleAsnIleValLysMetAlaIleM 629
QY 6338 TGCCCAAGCTAAATATTACATTCAGTCTACCCCATCAAGCTACCATTCATTCCTTCA 6397
Db 629 etProLysValIleThrArgPheAsnAlaIleProIleLysLeuProMetThrPhePhe 649
QY 6398 CAGAATTGGAAAAACAACCTTTAAATTTTCATATGGAACCAAAAAAGAGCCACAGAGCC 6457
Db 649 hrcLeuLeuGluLysThrThrLeuLysPheIleThrAsnGln-LysArgAlaArgIleAla 668
QY 6458 AGACAATCTTAAGCAAAAAAGAAAGCTGGAGTATCATGCTACCTGACTTTAAACTA 6517
Db 669 LysThrIleLeuSerGlnLysAsnLysAlaGlyIleThrLeuProAspPheLysLeu 688
QY 6518 TACTATAAGGCTACAGTAACCAAACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6577
Db 689 TyrThrLysAlaThrValThrLysThrAlaThrTrpThrTrpThrGlnAsnArgAspIleAsp 708
QY 6578 CAATGGAACAGACAGACCTCAGAAATTTAC-ACTGCAATCTACATCCATCTGACTTT 6636
Db 709 GlnTrpAsnArgIleGluProLeuGluLeuProHisIleThrAsnHisLeuIlePhe 728
QY 6637 GACAAACCTGACAAAAACAAGCAATGGAAGAGTTCCTATTTAAATAAATGGTGTGG 6696
Db 729 AspLysProAspLysAsnLysLeuTrpGlyLysAspSerLeuPheAsnLysTrpCysTrp 748
QY 6697 AAAAAGCTGCTAGCCATATGCAGAAAGCTGAAAGCTGATCCCTTCCTTACACCTTATA 6756
Db 749 GluAsnTrpLeuAlaIleCysArgLysLeuLysLeuAsnLeuPheLeuThrProThrThr 768
QY 6757 AAAGTTAACTCAAGATGAATTAAGACTTAAATATAAGACATAAAACCAATAAAACCCA 6815
Db 769 LysIleAspSerArgTrpIleLysAspLeuAsnValArgProLysThrIleLysIleLeu 788
QY 6816 GAAGAAACCTAGGCAATACCATTCAGGATATGGACATGGGCAAGACTTCATGACTAA 6875
Db 789 GluLysAsnLeuGlyAsnThrIleGlnAspIleGlyValGlyLysAspPheMetThrLys 808
QY 6876 ACACCAAAAGCAATGCAACAAAGCCAAATAGACAAAGTGGATCTGATTAACTATAG 6935
Db 809 ThrProLysAlaMetAlaThrLysAlaLysIleAspLysTrpAspIleLysLysLys 828
QY 6936 AGCTTCTGCACAGCAAAAAAACTGCTACAGTGAACAACCAACCTACAGAATGGGA 6995
Db 829 SerPheCysThrAla-LysGluThrThrIleIleValAsnArgGlnProThrGluTrpG 848
QY 6996 GAAATTTTGCATATCTGATCTGACAAAGGCTAATATCCAGAGATCTACAGAAGACT 7055
Db 848 uLysIlePheLysIleThrProSerAspLysGlyLeuIleSerArgIleThrLysGluLe 868
QY 7056 TAAACAATTTACAGAAAAA ---AACAAACCCGTCAAATATGGGCAAGGATATAG 7111
Db 868 uLysGlnIleThrLysLysSerAsnAsnProIleLysAsnTrpAlaLysAspMetAs 888
QY 7112 CAGACACTTCTCAAAAAGACATTTATGCAGCCCAACAACATATATCAAAAAACCCATC 7171
Db 888 nargHisPheSerLysGluAspIleThrAlaValAsnArgHisMetLysThrCysSerSe 908
QY 7172 ATCATTTGGTGTAGAGAAATGCAAAAAACCAACAGTGCATACCATCTCATGCTAGT 7231
```

```
Db 908 rLeuLeuAlaIleArgGluMetGlnIleLysThrMetArgThrHisPheThrProva 928
QY 7232 TAGAATGGTGATCACTAAATAAGTCAGAAACAACA 7268
Db 928 lArgMetAlaSerIleLysLysSerGlyAsnAsnArg 940

RESULT 6
US-10-025-201-3
; Sequence 3, Application US/10025201
; Publication No. US20030003468A1
; GENERAL INFORMATION:
; APPLICANT: Citow, Mary K.
; TITLE OF INVENTION: MARKERS FOR DISEASE SUSCEPTIBILITY AND TARGETS FOR THERAPY
; FILE REFERENCE: 5983/2H567
; CURRENT APPLICATION NUMBER: US/10/025, 201
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/256, 673
; PRIOR FILING DATE: 2000-12-19
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 1275
; TYPE: PRT.
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: GenBank Accession No. US20030003468A1 U09116
; DATABASE ENTRY DATE: 1995-02-02
; RELEVANT RESIDUES: (1)..(1275)
US-10-025-201-3

Alignment Scores:
Pred. No.: 8,79e-171 Length: 1275
Score: 2215.00 Matches: 408
Percent Similarity: 95.01% Conservative: 11
Best Local Similarity: 92.52% Mismatches: 22
Query Match: 4.23% Indels: 0
DB: 9 Gaps: 0

US-10-083-853B-2 (1-29921) x US-10-025-201-3 (1-1275)
QY 29919 ACTTCTTCCAGAAATGGAAAACTACTTTAAAGTTCATGTTGGAACCAAAAAAGAGCC 29860
Db 835 ThrPhePheThrGluLeuGluLysThrLeuLysPheIleThrLeuLysAsnGlnLysArgAla 854
QY 29859 CACATCGCAAGTCAATCTCTAAGCAAAAGAAAGAGAGAGTGGAGCATCATGCTACCTGAC 29800
Db 855 ArgIleAlaLysSerIleLeuSerGlnLysAsnLysAlaGlyIleThrLeuProThr 874
QY 29799 TTCAAACTACACTACAGGCTACAGTAACCAAAACAACATGGTACTGGTACCAAAAACAGA 29740
Db 875 PheLysLeuThrThrLysAlaThrValThrLysThrAlaThrTrpThrTrpThrGlnAsnArg 894
QY 29739 GATATCAACCAATGGAAACAGACAGAGCCCTCAGAATAATATCGCGCATATCTACACTAT 29680
Db 895 AspIleAspGlnTrpAsnArgThrGluProSerGluIleMetProHisIleThrAsnThr 914
QY 29679 CTGATCTTGCACAAACCTGAGAAAAACAAGCAATGGGAAGGATTCCTTATTATAATAA 29620
Db 915 LeuIlePheAspLysProGluLysAsnLysGlnTrpGlyLysAspSerLeuPheAsnLys 934
QY 29619 TGGTGTGAGAAAACTGGCTAGCCCTATGTAGAAAGCTGAAAGCTGAACTGCTTCTTACATA 29560
Db 935 TrpCysTrpGluAsnTrpLeuAlaIleCysArgLysLeuLysLeuAspLeuPheLeuThr 954
QY 29559 CCTATACAAAATAATTAATCAAGATGGATTAAAGACTTAAGATTTAGACTTAAACCATTA 29500
Db 955 ProThrThrLysIleAsnSerArgTrpIleLysAspLeuAsnValLysProLysThrIle 974
QY 29499 AATCCCTAGAGAAAAACCCAGCAATACCATTCAGGACATAGGATGGCAGGAGGACTTC 29440
Db 975 LysThrLeuGluGluAsnLeuGlyIleThrIleGlnAspIleGlyValGlyLysAspPhe 994
```

QY 29439 ATGCTTAAACACCAAGCAATGGCAACAAAGCCAAATTCAGAAATGGGATTAATT 29380
Db 995 MetSerLysThrProLysAlaMetAlaThrLysAspLysIleAspLysTrpAspLeuIle 1014
QY 29379 AAACCTAAAGAGCTTCTGCACACCAAGAAATACCATCAGAGTCAACAGGCAACTTACA 29320
Db 1015 LysLeuLysSerPheCysThrAlaLysGluThrThrIleArgValAsnArgGlnProThr 1034
QY 29319 GAATGGGAGAAATTTTGCACACCTACTCATTTGACAAAGGGCTAATATCCAGATCTAC 29260
Db 1035 ThrTrpGluLysIlePheAlaThrTrpSerSerAspLysGlyLeuIleSerArgIleTrp 1054
QY 29259 AATGAACCTAAACAAATTTTACAAGAAAGAAACCAAGACCCCATCAAAAGTGGGTGAAG 29200
Db 1055 AsnGluLeuLysGlnIleTrpLysLysLysThrAsnAsnProIleLysLysTrpAlaLys 1074
QY 29199 GATATGAACAGACACTTCTCAAGAAAGGCATTTATGAGCCCAAAACACATGAAAAA 29140
Db 1075 AspMetAsnArgHisPheSerLysGluAspIleTrpAlaLysLysHisMetLysLys 1094
QY 29139 TACTCATCATCACTGGCCATCAGAGAAATGCAAAATCAAAACCAACATGAGATACCATCTC 29080
Db 1095 CysSerSerSerLeuAlaIleArgGluMetGlnIleLysThrMetArgTrpHisLeu 1114
QY 29079 ACACGAGTTAGATGGCAATCAATTAAGAAAGTCAAGAAACACAGGTGGGAGAGATGT 29020
Db 1115 ThrProValArgMetAlaIleLysLysSerGlyAsnAsnArgCysTrpArgGlyCys 1134
QY 29019 GGAGAAATAGGAACACTTTTACACTGTTGGTGGGACTGTGAACCTAGTTCAACCATTTGG 28960
Db 1135 GlyGluIleGlyThrLeuLeuHisCysTrpTrpAspCysLysLeuValGlnProLeuTrp 1154
QY 28959 AAGTCAGTGTGTGATTCCTCAGGATCTAGAACTAGAAATACCAATTTGACCCAGCCATC 28900
Db 1155 LysSerValTrpArgPheLeuArgAspLeuGluLeuGluIleProPheAspProAlaIle 1174
QY 28899 CCATTACTGGGTATATACCCAAAGGATTATAATCATGCTGCTATAAGACACAAAGCACA 28840
Db 1175 ProLeuLeuGlyIleTrpProGluAspTrpLysSerCysCysTrpLysAspThrCysThr 1194
QY 28839 TGTATGTTATAGCAGCACTATTCACATAGCAAGACTTGAACCAACCTAAATGTCCA 28780
Db 1195 ArgMetPheIleAlaLeuPheThrIleAlaLysThrTrpAsnGlnProLysCysPro 1214
QY 28779 ACAACGATAGCTGATTAAGAAATGTGGCACATATACACCATGGAATACTATCGACC 28720
Db 1215 ThrMetIleAspTrpIleLysLysMetTrpHisIleTrpMetGluTrpTyAlaAla 1234
QY 28719 ATAAAAATGATGAGTTCGTGCTTCCTTTAGGACATGGATGAAGCTGGAACCAACATCAT 28660
Db 1235 IleLysAsnAspGluPheIleSerPheValGlyThrTrpMetLysLeuGluThrIleIle 1254
QY 28659 CTCAGCAACTATCACAGGACAAAGAAACCAACACCCGATGTTCTCACTCATGATGGGG 28600
Db 1255 LeuSerLysLeuSerGlnGluGlnLysThrLysHisArgIlePheSerLeuIleGlyIle 1274
QY 28599 AAT 28597
Db 1275 Asn 1275

RESULT 7

US-10-050-882-80
; Sequence 80, Application US/10050882
; Publication No. US20030104400A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 27 Human secreted proteins
; FILE REFERENCE: PZ038P1
; CURRENT APPLICATION NUMBER: US/10/050,882
; CURRENT FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: 09/661,453
; PRIOR FILING DATE: 2000-09-13
; PRIOR APPLICATION NUMBER: PCT/US00/06783

; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/125,055
; PRIOR FILING DATE: 1999-03-18
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 80
; LENGTH: 1010
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (25)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (104)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (194)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (362)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (525)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (643)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (649)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (656)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (660)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (731)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (770)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (777)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (790)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (800)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (825)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (987)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (996)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (1003)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-050-882-80

Alignment Scores:

Pred. No.: 1.57e-169 Length: 1010
Score: 2199.00 Matches: 405
Percent Similarity: 93.42% Conservative: 7
Best Local Similarity: 91.84% Mismatches: 29
Query Match: 4.20% Indels: 0
DB: 9 Gaps: 0

US-10-083-853B-2 (1-29921) x US-10-050-882-80 (1-1010)

QY 29919 ACATTTCTTCACAGAAATGGAAGAACTACTTTAAAGTTTCATGTGGAAACCAAAAAAGAGCC 29860
Db 570 ThrPhePheThrGluLeuGluLysThrThrLeuLysPheIleThrLeuLysPheIleThrLeuLysArgAla 589
QY 29859 CACATGCCCAAGTCAATCCTAAGCCAAAGAAAGAAAGCTGGAGGCATCATGCTACCTGAC 29800
Db 590 ArgIleAlaLysSerIleLeuSerGlnLysAsnLysAlaGlyGlyIleThrLeuProAsp 609
QY 29799 TTCAAACTACATACAGGCTACAGTAACCAAAACAAACATGTGTACCTACCAAAACAGA 29740
Db 610 PheLysLeuTyrTyrLysAlaThrValThrLysThrAlaTrpTyrTrpTyrGlnAsnArg 629
QY 29739 GATATCAACCAATGGACAGACAGAGCCCTCAGAAATATCCGCATATCTACACTAT 29680
Db 630 AspileAspGlnTrpAsnArgThrGluProSerGluIle***ProHisIleTyrAsn*** 649
QY 29679 CTGATCTTTGACAAACCTGAGAAACAAAGCAATGGGAAAGATTCCCTATTATAAAA 29620
Db 650 LeuIlePheAspLysPro***LysAsnLys***TrpGlyLysAspSerLeuPheAsnLys 669
QY 29619 TGTGCTGAGAAACTGGCTAGCCCTATGTAGAAAGCTGAAAGTGGATCCCTTCCTTACA 29560
Db 670 TrpCysTrpGluAsnTrpLeuAlaIleCysArgLysLeuLysLeuAspProPheLeuThr 689
QY 29559 CCTTATACAAAATTAATCAAGATGATTAAGACTTAAATGTTAGACCTAAACCATTA 29500
Db 690 ProTyrThrLysIleAsnSerArgTrpIleLysAspLeuAsnValArgProLysThrIle 709
QY 29499 AAATCCTTGAAGAAACCCAGCAATACCATTCAGACATAGGCATGGCAGGAGCTTC 29440
Db 710 LysThrLeuGluLysLeuGlyAsnThrIleGlnAspIleGlyMetGlyLysAspPhe 729
QY 29439 ATGCTTAACACCAAGCAATGGCAACAAAGCCAAATGACAAATGGGATCTAAT 29380
Db 730 Met***LysThrProLysAlaMetAlaThrLysAlaLysIleAspLysTrpAspLeuIle 749
QY 29379 AAATAAGAGCTTCGACAGCAAGAAAGAACTACCATCAGAGTGAACAGGCAACTTACA 29320
Db 750 LysLeuLysSerPheCysThrAlaLysGluThrIleArgValAsnArgGlnProThr 769
QY 29319 GAATGGAGAAATTTTTCACCTACTCATTTGACAAAGGGCTAATATCCAGAATCTAC 29260
Db 770 ***TrpGluLysIlePheAla***TyrSerSerAspLysGlyLeuIleSerArgIleTyr 789
QY 29259 AATGAATCAACAAATTTACAGAAGAAACAAAGAACCCCATCAAAAGTGGGTGAAG 29200
Db 790 ***GluLeuLysGlnIleTyrLysLysLys***AsnAsnProIleLysLysTrpAlaLys 809
QY 29199 GATATCAACAGACACTTCACAAAGAGGCATTTATGACAGCCAAAGAAACACATGAAAAA 29140
Db 810 AspMetAsnArgHisPheSerLysGluAspIleTyrAlaAlaLys***HisMetLysLys 829
QY 29139 TACTCATCATCTGCGCCATCAGAGAAATGCAAAATCAAAACCAATGAGATACCATCTC 29080
Db 830 CysSerSerLeuAlaIleArgGluMetGlnIleLysThrThrMetArgTyrHisLeu 849
QY 29079 ACACGAGTTAGATGCCATCATTTAAAGTACAGGAACACACAGGTGCTGGAGAGATGT 29020
Db 850 ThrProValArgMetAlaIleLysLysSerGlyAsnAsnArgCysTrpArgGlyCys 869
QY 29019 GGAGAAATAGGAACACTTTTACACTGTTGGTGGAGCTGTGAATCTTCAACCATTTGG 28960
Db 870 GlyGluIleClyThrLeuLeuHisCysTrpTrpAspCysLysLeuValGlnProLeuTrp 889
QY 28959 AAGTCAGTGTGTGATCTTCAGGGATCTAGAACTAGAAATACCAATTTGACCCAGCCATC 28900
Db 890 LysSerValTrpArgPheLeuArgAspLeuGluLeuGluIleProPheAspProAlaIle 909
QY 28899 CCATTTACTGGGTATATACCAAGGATTATAATCATGTGCTTATTAAGACACACAGCACA 28840
Db 899 LysSerValTrpArgPheLeuArgAspLeuGluLeuGluIleProPheAspProAlaIle 909

Db 910 ProLeuLeuGlyIleTyrProLysAspTyrLysSerCysCysTyrLysAspThrCysThr 929
QY 29839 TGTATGTTTATACGACACATATTCACAATAGCAAAAGACTTGGAAACCACTTAATGTCCA 28780
Db 930 ArgMetPheIleAlaAlaLeuPheThrIleAlaLysThrTrpAsnGlnProLysCysPro 949
QY 28779 ACACCATGAGACTGGGATTAAGAAATGTGGCACATATACACCATGAATACTATGCAGCC 28720
Db 950 ThrMetIleAspTrpIleLysLysMetTrpHisIleTyrThrMetGluTyrTyrAlaAla 969
QY 28719 ATAAATAATGATGAGTTCGTGCTCTTTAGGACATGGATGAAGCTGGAACCATCATTT 28660
Db 970 IleLysAsnAspGluPheMetSerPheValGlyThrTrpMetLysLeuGlu***IleIle 989
QY 28659 CTCAGCAAACTATCAAGACGACAAACCAACACGCGCATGCTCTCACTCATAGTGGG 28600
Db 990 LeuSerLysLeuSerGln***GlnLysThrLysHisArg***PheSerLeuIleGlyGly 1009
QY 28599 AAT 28597
Db 1010 Asn 1010

RESULT 8
US-10-000-256A-153
; Sequence 153, Application US/10000256A
; Publication No. US20030039983A1
; GENERAL INFORMATION:
; APPLICANT: Sun, Yongming
; APPLICANT: Recipon, Hervé
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Prostate Specific Genes a
; FILE REFERENCE: DEX-0259
; CURRENT APPLICATION NUMBER: US/10/000, 256A
; CURRENT FILING DATE: 2001-11-01
; PRIOR APPLICATION NUMBER: 60/244,782
; PRIOR FILING DATE: 2000-11-01
; NUMBER OF SEQ ID NOS: 240
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 153
; LENGTH: 956
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-000-256A-153

Alignment Scores:
Pred. No.: 1,45e-168 Length: 956
Score: 2187.00 Matches: 401
Percent Similarity: 94.51% Conservative: 12
Best Local Similarity: 91.76% Mismatches: 24
Query Match: 4.17% Indels: 0
DB: 9 Gaps: 0

US-10-083-853B-2 (1-29921) x US-10-000-256A-153 (1-956)

QY 29907 GAATTTGAAAAAATACTACTTTAAAGTTTCATGTGGAAACCAAAAGAGCCCATCGCAAG 29848
Db 520 GluLeuGluLysThrThrLeuLysPheIleTrpAsnGlnLysArgAlaHisIleAlaLys 539
QY 29847 TCATCTTAAGCCAAAGAAACAAAGCTGGAGGCATCATGCTACTGACTTCAAACTACAC 29788
Db 540 SerIleLeuAsnGlnLysAsnLysAlaGlyIleThrLeuProAspPheLysLeuTyr 559
QY 29787 TACAAGCTACAGTACCAACCAACATGGTACTGTGTACCAACAGAGATATCAACCA 29728
Db 560 TyrLysAlaThrValThrLysThrAlaTrpTyrTrpTyrGlnAsnArgAspIleAspGln 579
QY 29727 TGAACAGAACAGAGCCCTCAGAAATTAATGCCCATATCTACAACATATCTGATCTTTGAC 29668
Db 580 TrpAsnArgThrGluProSerGluIleThrGlnHisIleTyrSerTyrLeuIlePheAsp 599
QY 29667 AAACCTGAGAAAAACCAAGCAATGGGAAAGGATTCCTATTTAATAAATGGTCTGAGAA 29608
Db 599 LysSerValTrpArgPheLeuArgAspLeuGluLeuGluIleProPheAspProAlaIle 909

```
Db 600 LysProGluLysAsnLysGlnTrpGlyLysAspSerLeuPheAsnLysTrpCysTrpGlu 619
QY 29607 AACTGGCTAGCCTATGTAGAAAGCTGAAAGCTGAACTGCTCCCTTACACCTTATACAAAA 29548
Db 620 AsnTrpLeuAlaIleCysArgLysLeuLysLeuAspProPheLeuThrProTyrThrLys 639
QY 29547 ATTAATCAAGATGATTAAAGACTTAATGTAGACCTAAACCAATAAATCCCTAGAA 29488
Db 640 MetAsnSerArgTrpIleLysAspLeuAsnValArgProLysThrIleLysThrLeuGlu 659
QY 29487 GAAACCCAGGCAATACCATTCAGGACATAGCATGGCGAGGACTTCATGCTCTAAACA 29428
Db 660 GluAsnLeuGlyIleThrIleGlnAspIleGlyMetGlyLysAspPheMetSerLysThr 679
QY 29427 CCAAAAGCAATGGCAACAAAGCCAAATTTGACAATGGGATCTAATTAATAACTAAAGAGC 29368
Db 680 ProLysAlaMetAlaThrLysAspLysIleAspLysTrpAspLeuValLysLeuLysSer 699
QY 29367 TTCTGCACAGCAAAAGAACTACCATCAGAGTGACAGGCAACTTACAGAACTGGGAGAAA 29308
Db 700 PheCysThrAlaLysGluThrThrIleArgValAsnArgGlnProThrLysTrpGluLys 719
QY 29307 ATTTTGTCAACTACTCATTTGACAAAGGCTAATATCCAGAACTACAATGAACCTCAA 29248
Db 720 IlePheAlaThrTyrSerAspLysGlyLeuIleSerArgIleTyrAsnGluLysLys 739
QY 29247 CAAATTTCAAGAAAGAAACCAAGAACCCCATCAAAAGTGGTGCAAGGATATGACAGA 29188
Db 740 GlnIleTyrLysLysThrAsnAsnProIleLysLysTrpAlaLysAspMetAsnArg 759
QY 29187 CACTTCTCAAAGAGGCAATTTATGACGCCAAACACATGCAAAAAATATCATCATCA 29128
Db 760 HisPheSerLysGluAspIleTyrAlaAlaLysLysHisMetLysLysCysSerSer 779
QY 29127 CTGGCCATCAGAGAATCAATCAAAACCAATGAGATACCATCTCACACCAGTTAGA 29068
Db 780 LeuAlaIleArgGluMetGlnIleLysThrThrMetArgTyrHisLeuThrProValArg 799
QY 29067 ATGCAATCATTAAGCACTAGGAAACACAGGCTGCTGGAGAGATGTGGAGAAATAGGA 29008
Db 800 MetAlaIleLysLysSerGlyAsnAsnArgCysTrpArgGlyCysGlyGluThrGly 819
QY 29007 ACACHTTTACACTGTTGTGGGACTGTGAACCTAGTTCAACCATGTGTGGAAGTCAGTGG 28948
Db 820 ThrLeuLeuHisCysTrpTrpAspCysLysLeuAlaGlnProLeuTrpLysSerValTrp 839
QY 28947 TGATTCCTCAGGGATCTAGAACTAGAAATACCATTTGACCAGCATCCCATTTAGGGT 28888
Db 840 ArgPheLeuArgAspLeuGluLeuGluIleProPheAspProAlaIleProLeuLeuGly 859
QY 28887 ATATCCCAAGAGGATTATAATCATGCTGCTATAAAGACACAAAGCACATGTATGTTTATA 28828
Db 860 IleTyrProLysAspTyrLysSerCysCysTyrLysAspThrCysThrArgMetPheIle 879
QY 28827 GCAGCACTATTACATAGCAAGACTTGGAAACCACTAATGTCCAAACCAAGATAGAC 28768
Db 880 AlaAlaLeuPheThrIleAlaLysThrTrpAsnGlnProLysCysProThrIleIleAsp 899
QY 28767 TGGATTAAAGAAATGTGGACATATACACCATGGAATACTATGAGCCATATAAAATGAT 28708
Db 900 TrpIleLysLysMetTrpPheIleTyrThrMetGluTyrTyrAlaAlaIleLysAsnAsp 919
QY 28707 GAGTTCTGCTCTTTGTAGGACATGAGTGAAGCTGGAAACCATCATCTTCAGCAAACTA 28648
Db 920 GluPheValSerPheValGlyThrTrpMetLysLeuGluIleIleLeuSerLysLeu 939
QY 28647 TCACAGGACAAAAACCAACCCGATGTTCTCACTCATAGGTGGGAAT 28597
Db 940 SerGlnGluGlnLysThrHisArgIlePheSerLeuIleGlyGlyAsn 956
RESULT 9
US-10-083-853b-2
; Sequence 189, Application US/10001835
```

```

; Patent No. US20020160387A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Caferkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Ovary Specific Genes and
; FILE REFERENCE: DEX-0277
; CURRENT APPLICATION NUMBER: US/10/001,835
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/249,997
; PRIOR FILING DATE: 2000-11-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 189
; LENGTH: 917
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-001-835-189

Alignment Scores:
Pred. No.: 5.58e-110 Length: 917
Score: 1467.00 Matches: 275
Percent Similarity: 95.58% Conservative: 6
Best Local Similarity: 93.54% Mismatches: 13
Query Match: 2.80% Indels: 0
DB: 9 Gaps: 0

US-10-083-853b-2 (1-29921) x US-10-001-835-189 (1-917)
QY 29919 ACTTTCTTCACAGAAATGGAATAAAGTCTATGTAAGTCTGTAACCAAAAGAGGCC 29860
Db 624 ThrPhePheThrGluLeuGluLysThrThrLeuLysPheIleTrpAsnGlnLysArgAla 643
QY 29859 CACATCGCCAAAGTCAATCTTAAGCCAAAGAACCAAGCTGGAGGATCATCTACCTGAC 29800
Db 644 ArgIleAlaLysSerIleLeuSerGlnLysAsnLysAlaGlyIleThrLeuProAsp 663
QY 29799 TTCAAACTTACACTCAAGGCTACAGTACCAAAACACATGCTACTGTGTACCAAAACAGA 29740
Db 664 PheLysLeuTyrTyrLysAlaThrValThrLysThrAlaTrpTyrTrpTyrGlnAsnArg 683
QY 29739 GATATCAACCAATGGAACAGACAGCCCTCAGAAAATAATGCCGCATATCTACAACTAT 29680
Db 684 AspIleAspGlnTrpAsnArgThrGluProSerGluIleMetProHisIleTyrAsnTyr 703
QY 29679 CTGATCTTTGACAAACCTGAGAAAACCAAGCAANTGGGAAAGGATTCCTATTTAATAAA 29620
Db 704 LeuIlePheAspLysProGluLysAsnLysGlnTrpGlyLysAspSerLeuPheAsnLys 723
QY 29619 TGGTGTGAGAAAAGCTGGCTAGCCCTATGTAGAAAGCTGAAACCTGGATCCCTCTTACA 29560
Db 724 TrpCysTrpGluAsnTrpLeuAlaIleCysArgLysLeuLysLeuAspProPheLeuThr 743
QY 29559 CCTTATACAAAATAATTAATTCAGATGGAATTAAGACTTAAATGTTAGACCTTAAACCCATA 29500
Db 744 ProTyrThrLysIleAsnSerArgTrpIleLysAspLeuAsnValArgProLysThrIle 763
QY 29499 AAATCCCTAGAGAAAACCCAGGCAATACCATTCAGGCATAGGATGGGCAAGGACTTC 29440
Db 764 LysThrLeuGluGluAsnLeuGlyIleThrIleGlnAspIleGlyValGlyLysAspPhe 783
QY 29439 ATGCTTAAACACCAAAAGCAATGGCAACAAAGCCAAATTTGACAAATGGGATCTAAT 29380
Db 784 MetSerLysThrProLysAlaMetAlaThrLysAlaLysIleAspLysTrpAspLeuIle 803
QY 29379 AAATAAAGAGCTTCTGCACAGCAAAAGAACTACCATCAGATGCAACAGGCACTTACA 29320
Db 804 LysLeuLysSerPheCysThrAlaLysGluThrThrIleArgValAsnArgGlnProThr 823
QY 29319 GAATGGGAGAAAATAATTTTGGCAACCTACTCATTTGCACAAAGGGCTAATATCCAGATCTAC 29260
```

Db 824 ThrTrpGluLysIlePheAlaThrTyrSerAspLysGlyLeuIleSerArgIleTyr 843
QY 29259 AATGAACCTCAACAAATTTACAGAAACAAACAAAGACCCCATCAAAAGTGGTGAAG 29200
Db 844 AsnGluLeuLysGlnIleTyrLysLysThrAsnAsnProIleLysLysTyrPalaLys 863
QY 29199 GATATGAACAGACACTTCTCAAAAGAGGCAATTTATGACGCCCAAAACACATGAAACAAA 29140
Db 864 AspMetAsnArgHisPheSerLysGluAspIleTyrAlaAlaLysLysHisMetLysLys 883
QY 29139 TACTCATCATCTGGCCATCAGAGAAATCAAAATCAAAACCAATGAGATACCATCTC 29080
Db 884 CysSerSerSerLeuAlaIleArgGluMetGlnIleLysThrThrMetArgTyrHisLeu 903
QY 29079 ACACCACTTGAATGGCAATCATAAAAGCTCAGGAACAAAC 29038
Db 904 ThrProValArgMetAlaIleIleLysLysSerGlyAsnAsn 917
RESULT 10
US-10-082-830-235
; Sequence 235, Application US/10082830
; Publication No. US20030077604A1
; GENERAL INFORMATION:
; APPLICANT: Sun, Yongming
; APPLICANT: Recipon, Hervé
; APPLICANT: Salceda, Susana
; APPLICANT: Liu, Chenghua
; APPLICANT: Turner, Leah
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific
; FILE OF INVENTION: Genes and Proteins
; FILE REFERENCE: DEX-0249
; CURRENT APPLICATION NUMBER: US/10/082,830
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/243,802
; PRIOR FILING DATE: 2000-10-27
; NUMBER OF SEQ ID NOS: 282
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 235
; LENGTH: 940
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-082-830-235
Alignment Scores:
Pred. No.: 1.56e-100 Length: 940
Score: 1351.00 Matches: 251
Percent Similarity: 90.51% Conservative: 16
Best Local Similarity: 85.08% Mismatches: 28
Query Match: 2.58% Indels: 0
DB: 9 Gaps: 0
US-10-083-853B-2 (1-29921) x US-10-082-830-235 (1-940)
QY 29919 ACTTCTTCACAGAAATGGAAAACTACTTTAAAGTTCAATGTGGAACCAAAAGAGCC 29860
Db 646 ThrPhePheThrGluLeuGluLysThrThrLeuLysPheIleTrpAsnGlnLysArgAla 665
QY 29859 CACATCCCAAGTCATCTTAAGCCAAACAAAGACCTGGAGGCATCATGCTACCTGAC 29800
Db 666 ArgIleAlaLysThrIleLeuSerGlnLysAsnLysAlaGlyIleThrLeuProAsp 685
QY 29799 TTCAAACTACACTACAGGCTACAGTACCAACCAACATCGGTACTGTACCAAAACAGA 29740
Db 686 PheLysLeuTyrTyrLysAlaThrValThrLysThrAlaTrpTyrTrpTyrGlnAsnArg 705
QY 29739 GATATCAACCAATGGACAGACAGACCCCTCAAAATAATGCCGATATCTCAACACT 29680
Db 706 AspileAspGlnTrpAsnArgIleGluProLeuGluLeuIleProHisIleTyrAsnHis 725
QY 29679 CTGATCTTTACAAACCTTGACAAAACAAACAGCAATGGGGAAGATTCCCTATTATATAA 29620
Db 726 LeuIlePheAspLysProAspLysAsnLysLeuTrpGlyLysAspSerLeuPheAsnLys 745

QY 29619 TGTGCTGAGAAAACCTGGCTAGCCCTATGTAGAAAGCTGAAACTGGATCCCTTCTTACA 29560
Db 746 TrpCysTrpGluAsnTrpIleuAlaIleCysArgLysLeuLysLeuAsnLeuPheLeuThr 765
QY 29559 CCTATACAAAATAATTAATCAAGATGGATTAAAGACTTAATATTGTAGACCTAAACCAT 29500
Db 766 ProTyrThrLysIleAsnSerArgTrpIleLysAspLeuAsnValArgProLysThrIle 785
QY 29499 AATCCCTAGAAAACCCAGGCAATACCATTCAGACATAGCATGGCGCAGGACTTC 29440
Db 786 LysIleLeuGluLysAsnLeuGlyAsnThrIleGlnAspIleGlyValGlyLysAspPhe 805
QY 29439 ATGCTTAAACACCAAAAGCAATGGCAACAAAGCCAAAATTCACAAATGGGATCTAAT 29380
Db 806 MetThrLysThrProLysAlaMetAlaThrLysAlaLysIleAspLysTrpAspIleIle 825
QY 29379 AAATAAGAGCTTCTGCACAGCAAAAGAACTACCATCAGATGTAACAGCAACTTACA 29320
Db 826 LysLeuLysSerPheCysThrAlaLysGluThrThrIleIleValAsnArgGlnProThr 845
QY 29319 GAATGGAGAAAATTTTTCACACCTACTCATTTTGCACAAAGGGCTTAATATCCAGAACTAC 29260
Db 846 GluTrpGluLysIlePheLysIleTyrProSerAspLysGlyLeuIleSerArgIleTyr 865
QY 29259 AATGAACCTCAACAAATTTTACAAAGAAAACCAAGAACCCCATCAAAAGTGGTGAAG 29200
Db 866 LysGluLeuLysGlnIleTyrLysLysLysSerAsnAsnProIleLysAsnTrpAlaLys 885
QY 29199 GATATGAACAGACACTTCTCAAAAGAGGCAATTTATGACGCCCAAAACACATGAAACAAA 29140
Db 886 AspMetAsnArgHisPheSerLysGluAspIleTyrAlaValAsnArgHisMetLysThr 905
QY 29139 TACTCATCATCTGGCCATCAGAGAAATCAAAAGGCAATCAAAACCAACATGAGATACCATCTC 29080
Db 906 CysSerSerLeuLeuAlaIleArgGluMetGlnIleLysThrThrMetArgTyrHisPhe 925
QY 29079 ACACCACTTGAATGGCAATCATAAAAGCTCAGGAACCAACAGG 29035
Db 926 ThrProValArgMetAlaSerIleLysLysSerGlyAsnAsnArg 940
RESULT 11
US-10-025-201-2
; Sequence 2, Application US/10025201
; Publication No. US20030003468A1
; GENERAL INFORMATION:
; APPLICANT: Crow, Mary K.
; TITLE OF INVENTION: MARKERS FOR DISEASE SUSCEPTIBILITY AND TARGETS FOR THERAPY
; FILE REFERENCE: 5983/2H567
; CURRENT APPLICATION NUMBER: US/10/025,201
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/256,673
; PRIOR FILING DATE: 2000-12-19
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 338
; TYPE: PRT
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: GenBank
; DATABASE ENTRY DATE: 1995-02-02
; RELEVANT RESIDUES: (1)..(338)
US-10-025-201-2
Alignment Scores:
Pred. No.: 1.48e-79 Length: 338
Score: 1090.50 Matches: 234
Percent Similarity: 79.18% Conservative: 36
Best Local Similarity: 68.62% Mismatches: 67
Query Match: 2.08% Indels: 6
DB: 9 Gaps: 1

US-10-083-853B-2 (1-29921) x US-10-025-201-2 (1-338)

```
QY 2343 GGGAAAAACCACTGTCAGAGAAACACCTGAAATTCCTCAAAACACCAAGTCTCTCTCAACCA 2402
Db 2 GlyLysLysGlnAsnArgLysThrGlyAsnSerLysThrGlnSerAlaSerProPro 21
QY 2403 AAGGATCACTCTCTCCAGCAAGGGAACCAACAGATGGAGATGAGTTTGGAGAA 2462
Db 22 LysGluArgSerSerProAlaThrGluGlnSerTrpMetGluAsnAspPheAspGlu 41
QY 2463 TTGACAGAAAGTCTGAGAGAGTGGGTATATACAACTCTCTCGAGCTAAAGAGCAT 2522
Db 42 LeuArgGluGluGlyPheArg-----SerAsnTyrSerGluLeuArgGluAsp 58
QY 2523 GTTCTAACCCCAATCAAGCAAGCTTAAGAACCTTGAAGAAAGTGTAGATGAATGCTAACT 2582
Db 59 IleGlnThrLysGlyGluValGluAsnPheGluLysAsnLeuGluGluCysIleThr 78
QY 2583 AGAATAATCAGTGTAGAGAAACATAAATGACCTGTAGCTGAGCTGAAACGCAAGACAA 2642
Db 79 ArgIleThrAsnThrGluLysCysLeuLysGluLeuMetGluLeuLysThrLysAlaArg 98
QY 2643 GAACCTTCATGAGCATACACAGCTTCAATACCCAAATCGATCAAGCAGACAGAGGATA 2702
Db 99 GluLeuArgGluGluCysArgSerLeuArgSerArgCysAspGlnLeuGluGluArgVal 118
QY 2703 TCAGTGATTGAAGTCAAAATAAATAAGAGTGAAGAGCAAGATTACAGAAAAAGA 2762
Db 119 SerAlaMetGluAspGluMetAsnGluMetLysArgGluGlyLysPheArgGluLysArg 138
QY 2763 GTGAAAGAAACAACAAGCCTCCAGAAATATGGGACTATGTGAAAAGACCAATCTA 2822
Db 139 IleLysArgAsnGluGlnSerLeuGlnGluIleTrpAspTyrValLysArgProAsnLeu 158
QY 2823 CATTTGATTGTTGCTCCCAAGTATGGGAGATGGAATCAAGTTGGAAACACTCTT 2882
Db 159 ArgLeuIleGlyValProGluSerAspValGluAsnGlyThrLysLeuGluAsnThrLeu 178
QY 2883 CAGGTATTATCCAGGAGAATTCCTCAT-CTATCAGGCGCAGGCAACATTCAAATTCAG 2941
Db 179 GlnAspIleIleGlnGluAsnPheProAsnLeuAlaArgGlnAlaAsnValGlnIleGln 198
QY 2942 GAATATGGAGAACCACTAAGATATCTCTCGAGAGAAACAATCCCAAGACACATAATC 3001
Db 199 GluIleGlnArgThrProGlnArgTyrSerSerArgArgAlaThrProArgHisIleIle 218
QY 3002 TTCAGATTCACCAAGTGTGAATGAAGGAAAAATGTTAAGGCGCAGGAGAGAGAGGT 3061
Db 219 ValArgPheThrLysValGluMetLysGluLysMetLeuArgAlaAlaArgGluLysGly 238
QY 3062 TGGGTTACCCCAAGGAGCCCAATCAGACTAATCAGCGGATCTCCCGCAGAAACCCCTA 3121
Db 239 ArgValThrLeuLysGlyLysProIleArgLeuThrAlaAspLeuSerAlaGluThrLeu 258
QY 3122 CAAGCCAGAGAGTGAAGGCGCCATATTCACATTCCTTAAGAAATAATTTCAACCC 3181
Db 259 GlnAlaArgGluTrpGlyProIlePheAsnIleLeuLysGluLysAsnPheGlnPro 278
QY 3182 AGAATTTTCATATCCAGGCAACCAAGCTTCTTAAGTGAAGGAGAAATAAATCTCTCTACA 3241
Db 279 ArgIleSerPheProAlaLysLeuSerPheIleSerGluGlyGluArgLysTyrPheThr 298
QY 3242 GAGAGCAAAATGCTGACAGATTTTGTACACCAGGCGCTTCCTTACAGAGCTCCTGAA 3301
Db 299 AspLysGlnMetLeuArgAsp-PheValThrThrArgProThrLeuLysGluLeuLeuLys 318
QY 3302 GGAAGCACCC-AACATGGAAGAGACACTGTTACAGCCACTGCAAAAACATCCCAATT 3360
Db 318 sGluAlaLeuAsnMetGluArgAsnAsnArgTyrGlnProLeuGlnAsnHisAlaLysme 338
QY 3361 G 3361
Db 338 t 338
```

RESULT 12

```
US-10-083-853-2
; Sequence 2, Application US/10083853
; Patent No. US20020164709A1
; GENERAL INFORMATION:
; APPLICANT: Affymetrix, Inc T
; APPLICANT: Shigeta, Ron T
; APPLICANT: Siani-Rose, Michael A
; TITLE OF INVENTION: Nucleic Acid Encoding Growth Factor Protein
; FILE REFERENCE: 3385.1
; CURRENT APPLICATION NUMBER: US/10/083.853
; CURRENT FILING DATE: 2002-02-26
; PRIOR APPLICATION NUMBER: USSN 60/272.663
; PRIOR FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 317
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-083-853-2
```

```
Alignment Scores:
Pred. No.: 2,04e-56 Length: 317
Score: 806.00 Matches: 182
Percent Similarity: 51.88% Conservative: 11
Best Local Similarity: 48.92% Mismatches: 25
Query Match: 1.54% Indels: 154
Db: 9 Gaps: 6
```

US-10-083-853B-2 (1-29921) x US-10-083-853-2 (1-317)

```
QY 29454 ATGGCAGGAGTTCATCTCTAAACACCAACCAATGGCAACAAAGCCCAAAATTTGAC 29395
Db 1 MetGlyLysAspPheMetSerLysThrProLysAlaMetAlaThrLysAlaLysIleasp 20
QY 29394 AAATGGGATCTTAATAACTAAAGAGCTTCTGCACAGCAAAAGAACTACCATCAGAGTG 29335
Db 21 LysTrpAspLeuIleLysLeuLysSerPheCysThrAlaLysGluThrThrIleArgVal 40
QY 29334 AACAGGCAACTTACAGATGGGAGAAATTTTTCACACCTACTCATTTGACAAAGGGCTA 29275
Db 41 AsnArgGlnLeuThrGluTrpGluLysIlePheAlaThrTyrSerPheAspLysGlyLeu 60
QY 29274 ATATCCAGAACTCTCAATGAACTCAACAAATTTTACAGAAAAAACAAGAACCCCATC 29215
Db 61 IleSerArgIleTyrAsnGluLeuLysGlnIleTyrLysLysLysThrLysAsnProIle 80
QY 29214 AAAAAGTGGGTGAAGGATATGAACAGACACTTCTCAAAGAGGCGATTTATGCGACCCAA 29155
Db 81 LysLysTrpValLysAspMetAsnArgHisPheSerLysGluGlyIleTyrAlaAlaLys 100
QY 29154 AACACATGAAAAAATCTCATCTACTGCGCCATCAGAGAAATGCAAAATCAAAACCCACA 29095
Db 101 LysHisMetLysLysTyrSerSerSerLeuAlaIleArgGluMetGlnIleLysThrThr 120
QY 29094 ATGAGATACCATCTCACACCACTAGTAGATGCGCAATCAATTAAGTACAGGAAACACAGG 29035
Db 121 MetArgTyrHisLeuThrProValArgMetAlaIleIleLysLysSerGlyAsn--Asn- 139
QY 29034 TGCTGGAGAGAATGTGGAGAAATAGGAACACTTTTACACTGTTGGTGGGACTGTGAAC 28975
Db 139 ----- 139
QY 28974 GTTCAACCATTTGTGAAGTCAAGTGTGGTGTATTCCTCAGGGATCTAGAACTAGAAATACCA 28915
Db 139 ----- 139
QY 28914 TTTGACCCAGCCATCCCATTTACTGGGTATATACCAAGGATTATAAATCATCTGCTGCTAT 28855
Db 139 ----- 139
```


[illegible]

RESULT 15
US-10-074-475-250
; Sequence 250, Application US/10074475

DEC 13 1961

Publication No. US20030092898A1

GENERAL INFORMATION:
APPLICANT: Salceda, Susana
APPLICANT: Macina, Roberto
APPLICANT: Hu, Ping
APPLICANT: Recipon, Herve
APPLICANT: Karra, Kalpana
APPLICANT: Cafferkey, Robert
APPLICANT: Sun, Yongming
APPLICANT: Liu, Chenghua
TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific
FILE OF INVENTION: Genes and Proteins
FILE REFERENCE: DEX-Q313
CURRENT APPLICATION NUMBER: US/10/074,475
CURRENT FILING DATE: 2002-02-13
PRIOR APPLICATION NUMBER: 60/268,292
PRIOR FILING DATE: 2001-02-13
NUMBER OF SEQ ID NOS: 295
SOFTWARE: PatentIn version 3.1
SEQ ID NO 250
LENGTH: 190
TYPE: PRT
ORGANISM: Homo sapien
US-10-074-475-250

Alignment Scores:
Pred. No.: 6e-46 Length: 190
Score: 676.00 Matches: 147
Percent Similarity: 82.65% Conservative: 15
Best Local Similarity: 75.00% Mismatches: 25
Query Match: 1.29% Indels: 10
DB: Gaps: 3

US-10-083-853B-2 (1-29921) x US-10-074-475-250 (1-190)

QY 4151 TTAGAACTCAGAATTAGAACTCCTCAAAATCACAACTACATGGAACACAC 4210
Db 3 LeuGlnLeuArgIleLysSerLeuThrGlnAsnArgThrThrTrpLysLeuAsn 22
QY 4211 CTGCTCTGAATCAGTACTGGTAAATAACAAATGAAGGCAAAATAAAGATGTTCTT 4270
Db 23 LeuLeuAsnAspTrpIrpValAsnLysLysIleLysAlaGluIleAsnLysPhePhe 42
QY 4271 GAAACCAATGAGAACACACAAATGTACAGAACTCTCGGGGCATATTAAAGCAGTG 4330
Db 43 GluThrIleGluAsnLysAspThrMetTyGlnAsn-----ThrAlaLysAlaVal 59
QY 4331 TGTAGAGGGAATTTATAGACTAGATGCTCAAGAGAAAGCAGGAATATCTAAATA 4390
Db 60 PheArgGlyLysPheIleAlaLeuAsnThrHisIleArgAsnTrpGluIleProLysIle 79
QY 4391 GACACCTTAACATCACAAATTAAGAACTAGAGAGAAAGACAAACAATTCAAAAGCT 4450
Db 80 AsnValLeuThrSerGlnLeuLysGluLeuLysArgGluGlnThrHisSerLys--- 98
QY 4451 AGCAGAAGACAGAAATAACTAAGATCAGAGCAAGCTGAAGGAGATAGAGACACAAAA 4510
Db 99 -----GlnGluIleThrLysIleIleAlaGluLeuLysGluIleGluThrGlnLys 115
QY 4511 GCCCTTCAATAATCAATGAATCCAGGAGCTGGTTTTTGAAGAGATCAGCAAAATAGA 4570
Db 116 AlaLeuGlnLys-IleSerAspSerArgSerTrpPheGluLysIleAsnLysThrAs 135
QY 4571 CCA---CTAGACAGACATAAAGAAAGAGAGAGCAATCAAGAGATGCAATAAA 4626
Db 135 PArgLeuLeuAlaArgIleIleLysLysLysArgGluLysAsnGluIleAspThrIleLys 155
QY 4627 AAATGATAAGGGGATATCACCACCGATCCACAGAAATACAAACTATTATCAGAGAATA 4686
Db 155 sasnAspLysGlyAspIleThrThrAsnProThrGluIleGlnThrAlaIleArgGluCy 175
QY 4687 TTATAAACCTCTATGCAAAATAAACTAGAAATCTAGAAGAAATG 4732
Db 175 sasnAspLysGlyAspIleThrThrAsnProThrGluIleGlnThrAlaIleArgGluCy 175

Db 175 sTyrGlnHisLeuTyrrileAsnLysLeuGluAsnLeuGluGluIle 190

Search completed: July 4, 2003, 23:32:11
Job time : 761.5 secs

PF 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

PA Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX N-PSDB; AAS66529.

XX New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity

PS Claim 20; SEQ ID NO 32701; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving

CC (II). (II) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. ABG00010-ABG30377 represent novel human

CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 2747 AA;

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
0	5266.50	2747	1108
Percent Similarity:	71.53%	Conservative:	108
Best Local Similarity:	65.18%	Mismatches:	182
Query Match:	10.06%	Indels:	309
DB:	22	Gaps:	10

US-10-083-853B-2 (1-29921) x ABG02342 (1-2747)

Qy	2403	AAGATCACAACTCCTGCCCAAGGAAGGAAACCAACAGATGGAGATGAGTTTGAGGAA	2462
Db	1154	GlulGluArgSerSerProAlaThrGluGlnSerThrGluAsnAspPheAspGlu	1173
Qy	2463	TTGACAGAGTAGGCTTCAGAGGTGGGTAAACAACTCCTCCGAGCTAAAGAGAGCAT	2522
Db	1174	LeuArgGluGluGlyPheArgArg-----SerAsnTyrSerGluLeuArgGluAsp	1190
Qy	2523	GTTCTAACCCAACTCAAGGAGCTAAGAACCTTGAAAGAGGTTAGATGATTCCTAACT	2582
Db	1191	IleGlnThrLysGlyLysGluValGluAsnPheGluLysAsnLeuGluGluCysIleThr	1210
Qy	2583	AGAATAATACGTGTAGAGAAACAATAAATACCTGATGGAGCTGAAAGACCAAGACAA	2642
Db	1211	ArgIleThrAsnThrGluLysCysLeuLysGluLeuMetGluLeuLysThrLysAlaArg	1230
Qy	2643	GAACCTCATGAGCATCACAGCTTCAATGACCAATCGCATCAAGCAGAGAAAGGATA	2702
Db	1231	GluLeuArgGluGluCysArgSerLeuArgSerGlnCysAspGlnLeuGluGluArgVal	1250

Qy	2703	TCACTGATTTGAAGATCAAAATTAATAAAGAAAGTGAAGACACAGATTACAGAAAAAGA	2762
Db	1251	SerAlaMetGluAspGluMetAsnGluMetLysGlnGluGlyLysPheArgGluLysArg	1270
Qy	2763	GTGAAAGAAACAAACAAGCCTCCAAGATTTATGGGACTATGTGAAAAGACCAATCTA	2822
Db	1271	IleLysArgAsnGluGlnSerLeuGlnGluIleTrpAspTyrValLysArgProAsnLeu	1290
Qy	2823	CATTGTTGTTGGTGTCCCCCAAGATGGGGAGAGATGAATCAAGTTTGGAAAACACTCTT	2882
Db	1291	ArgLeuIleGlyValProGluSerAspAlaGluAsnGlyThrLysLeuGluAsnThrLeu	1310
Qy	2883	CAGGTTATTTCCAGAGGAA--TTTCCCATCTATCAGGGCAGGCCAGCAATTCATTAATTCAG	2941
Db	1311	GlnAspIleIleGlnGluThrPheProAsnLeuAlaArgGlnAlaAsnValGlnIleGln	1330
Qy	2942	GAATATGGGAGACACACCAATAGATACCTCTCGAGAGAAACAATCCCAAGACACATAATC	3001
Db	1331	GluIleGlnArgThrProGlnArgTyrSerSerArgArgAlaThrProArgHisIleIle	1350
Qy	3002	TTCAGATTACCAAGTTTGAATGAAGGAAAAAATGTTAAGGGCAGCCAGAGAGAAAGGT	3061
Db	1351	ValArgPheThrLysValGluMetLysGluLysMetLeuArgAlaAlaArgGluLysGly	1370
Qy	3062	TGGGTTACCCCAAGGAAAGCCCAATCAGACTACAGCGGATCTCCCGGCAAGAACCTTA	3121
Db	1371	ArgValThrLeuLysGlyLysProIleArgLeuThrAlaAspLeuSerAlaGluThrLeu	1390
Qy	3122	CAAGCCAGAAGAGTAGTGGGCCCAATATCCACACTTCTTAAGAAAAATAATTTTCAAGCC	3181
Db	1391	GlnAlaArgArgGluTrpGlyProIleLeuAsnIleLeuLysGluLysAsnPheGlnPro	1410
Qy	3182	AGAATTTTCATATCCAGCCAAACCAAGCTTCTTAAGTGAAGGAGAAATAAATCTCTTACA	3241
Db	1411	ArgIleSerTyrProAlaLysLeuSerPheIleSerGluGlyGluIleLysTyrPheThr	1430
Qy	3242	GAGAGCAATGCTGACAGATTTTTCACACCAGCGCTCCCTTACAGAGCTCCTGAA	3301
Db	1431	AspLysGlnMetLeuThrAsp-PheValThrSerArgProAlaLeuLysGluLeuLeuLys	1450
Qy	3302	GGAAGACCC--AACATGGAAGGAACAACCTGTGTACCGCCACTGCAAAAAACATCCCAATT	3360
Db	1450	SGLuAlaLeuAsnMetGluArgAsnAsn-----	1459
Qy	3361	GTAAGACCATTTGATGCTATGAAGAAAGTGCATCACTAACGGGCAAAATTAACCAAGTAG	3420
Db	1459	-----	1459
Qy	3421	TGTCATAATGCGCAGGATCAAAATTCACACATAATAATTAACCTTAATGTAATGGGCT	3480
Db	1459	-----	1459
Qy	3481	AAATTTCCCAATTAAAGACACAGACTGGCAAAATTTGGATAAAGAGTCAAGACCCATCAGT	3540
Db	1459	-----	1459
Qy	3541	GTGCTGTATTTCAGGAGGCCCATCTCACATGAAAAGACACACATAGGCTCAAAATAAGGG	3600
Db	1460	-----ArgAspThrHisArgLeuLysIleLysGln	1469
Qy	3601	ATGAGGAGAGATTTTACCAAGTAAATGGAACCAAAAAAAGCAGGGGTTCGAATCCT	3660
Db	1469	YTPArgLysIleTyrGlnAlaAsnGlyLysGln---LysLysAlaGlyValAlaIleLe	1488
Qy	3661	AGTCTCTGATAAAACAGACTTTAAACCAACAAAGATCAAAAGACAAAAGAGCCCATTA	3720
Db	1488	uValSerAspLysThrAspPheLysProThrLysIleLysArgAspLysGluGlyHisTy	1508
Qy	3721	CATAATGGTAAAGGCATCAATGGAACAAAGAGACTTAATCTTAATATATACATGCACC	3780
Db	1508	rMetMetValLysGlySerIleGlnGluGluLeuThrIleLeuAsnMetTyrAlaPr	1528
Qy	3781	CAATACAGAGGAGACCCAGATTTCATAAAGCAAGTTCTTAGAGACCTACAAAGAGACTTTGA	3840

Db	1528	oAsnThrGlyAlaProArgPheIleLysGlnValLeuSerAspLeuGlnArgAspLeuAs	1548	1845	-----	1844
QY	3841	CTCCACACAATAATAGTGGAGGCTCTAAATAATAATATAGACACTTTAAACACCCACTGCC	3900	QY	4917	AGATAGCAGAGAGGGTGAGTTGGAAGAAGAGGTCAGCTGTTAGGTCACAGAGTAGGCCTGCA
Db	1548	pSerHisThrLeuIleMetGly-----AspPheAsnThrProLeuSe	1562	Db	1845	-----
QY	3901	AAATATTAGGCAGATCAATAGACAGACAGAAAATTAAACAGGATATATCCAGGAGTTGAAC	3960	QY	4977	AGACTGCATTCCTCAAAACAATAGGCTCTAGATGTCCCAAGTAGATAACCTCAAGAGGCCAG
Db	1562	rThrLeuAspArgSerThrArgGlnLysValAsnLysAspThrGlnGluLeuAsnSerAl	1582	Db	1845	-----
QY	3961	TCGGACACAGCGGACCTAATAGATATCTACAGAACTCCACCACCAATCAACAGAAATA	4020	QY	5037	TGCCAGGAGTGATGCCCTCAGCAAAACCTTCTAGGGCAGGCACACAGAACTTAAGTTTGCCC
Db	1582	aLeuHisGlnAlaAspLeuIleAspIleTyrArgThrLeuHisProLysSerThrGluTy	1602	Db	1845	-----
QY	4021	TACACTCTCTCAGCATCACATTACACCTATTTTAAATTTGACCATGTAAATTTTAAGTAA	4080	QY	5097	ACATTCTGTATTACGCATAAACAGTTTGCTGTTTGTATCAAGTAGCCTCCAGTCCGAATGCT
Db	1602	rThrPhePheSerAlaProHisHisSerTyrSerLysIleAspHisIleLeuGlySerGI	1622	Db	1845	-----
QY	4081	AACACTCCTCAGCAAAATGCAAAAGACAGAAAATCCTAAACAACAGTCTCTCAGACTACAG	4140	QY	5157	GAGTTGGTCATGATCCCTTTGGCCTTTTGGCTCCCAAACACACATACACCTCTCAAGAC
Db	1622	uAlaLeuLeuSerLysCysLysArgThrGluIleThrAsnTyrLeuSerAspHisSe	1642	Db	1846	-----
QY	4141	TGCAATCTATTAGAACTCAGAATTAGAACAATCACTCAAAATCACACACTACATGGAA	4200	QY	5217	TAAACCAGGAAGTCAAAATCCCTCAATATACCAGTAAACAAGTTCTTAAATTTGAAGCAG
Db	1642	rAlaIleLeuLeuGluLeuArgIleLysAsnLeuThrGlnSerArgSerThrTrpLy	1662	Db	1852	eUAsnGlnGluValGluSerLeuAsnArgProIleThrGlyAlaGluIleValAlaI
QY	4201	ACTGAACAACCTGCTCTGAATGACTACTGGGTAAATAACAAAATGAAGSCAAAATAAA	4260	QY	5277	TAATTGATGCTTACCACCAACAAAAAAGTCCAGGACACAGCGGATTCACAGCCAAATCT
Db	1662	sLeuAsnAsnLeuLeuLeuAsnAspTyrTrpValHisAsnGluMetLysAlaGluIleLy	1682	Db	1872	leIleAsnSerLeuSerThrLysLysSerProGlyProaspGlyPheThrAlaGluPheT
QY	4261	GATGTTCTTTGAACCAATGAGAACAGACACAATGTACCAGAAATCTCTGGGGCATATT	4320	QY	5337	ACCAGAGGTACAAAGAGAAGCTGGTACTATTCTCTGAAACTATTTCCAAAAAATAAGAA
Db	1682	sMetPhePheGluThrAsnGluAsnLysAspThrThrTyrGlnAsnLeuTrpAspAlaph	1702	Db	1892	yrGlnArgTyrLysGluLeuValProPhePheLeuLysLeuPheGlnSerIleGluL
QY	4321	TAAAGCACTGTGTAGAGGGAATTTATAGCACTAGATGCCTACAGAGAAGACAGAAAT	4380	QY	5396	--AATGGGAATCCTCCCTAACTCATTTTACGAGGCAGCATCATCTGTATACCAAAACCT
Db	1702	eLysAlaValCysArgGlyLysPheIleAlaLeuAsnAlaHisLysArgLysGlnGluAr	1722	Db	1912	ysGlu-GlyIleLeuProAsnSerPheTyrGluAlaSerIleIleLeuIleProLysPro
QY	4381	ATCTAAATATAGACACCTTAACATCACAANTTAAAGAACTAGAGAAGAAAGACAAACAA	4440	QY	5454	AGCAGTGCACACACAAAAAGAGAAATTTTCAGGCCCATATCCCTGATGAACATTGATGTG
Db	1722	gsrLysIleAspThrLeuThrSerGlnLeuLysGluLeuGluLysGlnGluInThrHI	1742	Db	1932	GlyArgAspThrThrLysLysGluAsnPheArgProIleSerLeuMetAsnIleAspAla
QY	4441	TTCAAAAGCTAGCAGAGACAGAAATACTAAGATCAGACGAACTCAAGAGATAGA	4500	QY	5514	AAATCCTCAATAAAATACTGGCAACCAAAATCCAGCAGACATCAAAAAGCTTATCTAC
Db	1742	sSerLysAlaSerArgArgGlnGluIleThrLysIleArgAlaGluLeuLysGluIleGI	1762	Db	1952	LysIleLeuAsnLysIleLeuAlaAsnArgIleGlnGlnHisIleLysLysLeuIleHis
QY	4501	GACAAAAAGCCCTCAATAATAATCAATCAATCCAGGAGCTGGTTTTTGAAGAATCA	4560	QY	5574	CATGATCAAGTTGGCCTCATCCTCGGATGCAAGGCTGCTTCAAAATATGCCAAATCAATA
Db	1762	uThrGlnLysThrLeuGlnLys-IleAsnGluSerArgSerTrpPhePheGluArgIlea	1782	Db	1972	HisAspGlnValGlyPheIleProGlyMetGlnGlyTrpPheAsnIleArgLysSerIle
QY	4561	GCAAAAT-AGACACTAGACACTAATAAAGAAAGAAAGACAGAGAATCAAAAGAG	4616	QY	5634	AATGTAGGCATCACATAAACAGAACCAATGACAAAAACCATGATTTATCTCAATAGAT
Db	1782	snLysIleAspArgProLeuAlaArgLeuIleLysLysLysArgGluLysAsnGlnIleA	1802	Db	1992	AsnValIleGlnHisIleAsnArgAlaLysAspLysAsnHisMetIleValSerIleAsp
QY	4617	ATGCAATAAAAATGATAAAGGGATATACCACCGATCCACAGAAATACAAACTATTA	4676	QY	5694	GCAGAAAAGCCTTTGCTCAAAATTCACAGCCCTTCATGCTAAAAATTCCTAGTAACATA
Db	1802	spThrIleLysAsnAspLysGlyAspIleThrThrAspProThrGluIleGlnThrI	1822	Db	2012	AlaGluLysAlaPheAspLysIleGlnGlnProPheMetLeuLysThrLeuAsnLysLeu
QY	4677	TCAGAGAAATATTATAACACCTCTATGCCAATAAATAGAGAAATCTAGAGAATATGATA	4736	QY	5754	GGTATCGATGGAATGTATCTCAAAATAATAAGAGCTATTTATACAAACCCACAGCCAATA
Db	1822	leArgGluTyrTyrLysHisLeuTyAlaAsnLysLeuGluAsnLeuGluGluMetAspt	1842	Db	2032	--Val-----
QY	4737	AATTCTCTGGACACATATGTAGCTGTATGACCTTGGGGGACAGACAAAAGGGGTGAA	4796	QY	5814	TCATACTGAATGGCAAAAACCTGGAAGCATTCCTTTGAGAACTGGCACAAAGCAAGGAT
Db	1842	hrPheLeuAsp-----	1845	Db	2032	-----
QY	4797	TGCAGAAATAAAAGACAAAAGACAAAAGAGTATGTTTGAAGTAGGGGTACAGGGGCAACT	4856	QY	5874	GCCCTCTCACCACTCCTATTCAAGATACTATTGGAAGTTCTTGGCCAGGCGCAATCAGCG
Db	1845	-----	1845	Db	2033	-----
QY	4857	TGCTCTTAATGGACAAGGCCCTTGAGCTTTACACCACCCCTCTGTATTTATTAGGCAAAAG	4916	QY	5934	AATAGAAAATAAAGGGTATTCAAAATAGAAAGAGGAAAGTATTTGCTCTCTGTTTG

XX

The polynucleotide sequences given in AAF33213 to AAF33261 encode the human secreted proteins given in AAB64882 to AAB64930. AAB64931 to AAB64991 represent human secreted polypeptide sequences and proteins homologous to them, which are given in the exemplification of the present invention. Human secreted proteins have activities based on the tissues and cells the genes are expressed in. Examples of activities include: immunomodulatory; antisclerotic; dermatological; immunosuppressive; anti-inflammatory; anti-HIV; immunostimulant; cytostatic; cardiant; vascular; antimicrobial; anti-angiogenic; ophthalmological; neuroprotectant; anticonvulsant; nootropic; antialzheimers; antiparkinsonian; and vulnerary. The polynucleotides and polypeptides can be used in the prevention, diagnosis and treatment of diseases associated with inappropriate polypeptide expression. Disorders that may be prevented, diagnosed and/or treated by the above methods include immune disorders (e.g. multiple sclerosis, systemic lupus erythematosus and human immuno-deficiency virus (HIV), infections), hyperproliferative disorders (e.g. cancers and Gaucher's disease), cardiovascular diseases (e.g. Scimitar syndrome, Chaga's cardiomyopathy and coronary arteriosclerosis), angiogenic disorders (e.g. corneal graft neovascularisation and diabetic retinopathy), neurological disorders (e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease), infectious diseases and/or for promoting wound healing, regeneration and/or chemotaxis. AAF33204 to AAF33212 and AAB64881 represent sequences used in the exemplification of the present invention.

XX

SQ Sequence 1280 AA:

Alignment Scores:
 Pred. No.: 0 Length: 1280
 Score: 5244.00 Matches: 1068
 Percent Similarity: 78.90% Conservative: 69
 Best Local Similarity: 74.12% Mismatches: 138
 Query Match: 10.02% Indels: 173
 DB: 22 Gaps: 7

US-10-083-853B-2 (1-2921) x AAB64943 (1-1280)

QY 3413 CCAGCTAGTGTGCATATGCGAGGATCAAAATTCACACATAATAATTAACCTTAATGTA 3472
 Db 1 ProAlaAsnIleIleMetThrGlySerAsnSerHisIleThrIleLeuThrLeuAsnVal 20
 QY 3473 ANTGGCTAAATCCCAATTAAGACACACAGCTGGCAAAATGGTAAGAGTCAAGAC 3532
 Db 21 AsnGlyLeuAsnAlaProIleLysArgHisArgLeuAlaAsnTrpIleLysSerGlnAsp 40
 QY 3533 CCATCAGTGTGTGTATTCAGGAGGCCCATCTCACATGAAGAAGACACACATAGGCTCAA 3592
 Db 41 ProSerValCysCysIleGlnGluThrHisLeuThrCysArgAspThrHisArgLeuLys 60
 QY 3593 ATAAAGGGATGGAGGAAGATTTACCAAGTAAATGGAACCAAAACAAAAAAGCAGGGTT 3652
 Db 61 IleLysGlyTrpArgIleIleThrGlnAlaAsnGlyLysGln---LysLysAlaGlyVal 79
 QY 3653 GCAATCCTAGTCTCTGATTAACAGACTTTAAACCAACAAAGATCAAAAGACAAAGAA 3712
 Db 80 AlaIleLeuValSerAspLysThrAspPheLysProThrLysIleLysArgAspLysGlu 99
 QY 3713 GGCATTACATATGTTAAAGGCATCAATGGACAGAGAGCTAACTATCTTAATATA 3772
 Db 100 GlyHisTrpIleMetValLysGlySerIleGlnGlnGluLeuThrIleLeuAsnIle 119
 QY 3773 CATGCACCAATACAGGAGCACCAGATTTCATAAGCAAGTCTTAGAGACCTACAAAGA 3832
 Db 120 TyrAlaProAsnThrGlyAlaProArgPheIleLysGlnValLeuSerAspLeuGlnArg 139
 QY 3833 GACTTTGACTCCCAACAATAATAGTGGAGTCTAAATAATAATAGACACTTTACACC 3892
 Db 140 AspLeuAspSerHisThr***IleMetGly-----AspPheAsnThr 153
 QY 3893 CCACCTGCCAATATTAGGCAGATCAATGACACAGAAAATTAACAGGATATCAGGAGTTG 3952
 Db 154 ProLeuSerThrLeuAspArgSerThrArgGlnLysValAsnLysAspThrGlnGluLeu 173

QY 3953 AACTGAGCTCTGACCAACGCGACCTAATAGATATCTACAGAACTCCCCACCCCAATCA 4012
 Db 174 AsnSerAlaLeuHisGlnAlaAspLeuIleAspIleTyrArgThrLeuHisProLysSer 193
 QY 4013 ACAGATAATACACTCTCTCAGCATCACAATACACCTATTTTAAATGACCATGTAAT 4072
 Db 194 ThrGluTyrThrPhePheSerAlaProHisHisThrTyrSerLysIleAspHisIle*** 213
 QY 4073 TTAAGTAAACACTCTCTCAGCAAAATGCCAAAGAACAAATCTCTACAACAGCTCTCTCA 4132
 Db 214 GlySerLysAlaLeuLeuSerLysCysLysArgThrGluIleIleThrAsn***LeuSer 233
 QY 4133 GACTACAGTGCATCTATTAGAACCTCAGAAATTAAGAAATCTACTCAAAATCACAACT 4192
 Db 234 AspHisSerAlaIleLysLeuGluLeuArgIleLys***LeuThrGlnAsnArgSerThr 253
 QY 4193 ACATGGAACCTGAACAACTGCTCTGTAATGACTACTGGGTAAATAACAAAATGAAGCA 4252
 Db 254 ThrTrpLysLeuAsnAsnLeuLeuLeuAsnAspTyrTrpValHisAsnGluMetLysAla 273
 QY 4253 AAAATAAAGATCTCTTTGAACCAATGAGAACAAAGACACATGTACCAGATCTCTGG 4312
 Db 274 GluIleLysMetPheGluThrAsnGluAsnLysAspThrThrTyrGlnAsnLeuTyrp 293
 QY 4313 GGCATATTTAAAGCAGTGTGTAGAGGGAATTTATAGCACTAGATGCCTACAGAGAAAAG 4372
 Db 294 Asp***PheLysAlaValCysArgGlyLysPheIleAlaLeuAsnAlaHisLysArgLys 313
 QY 4373 CAGGAATATCTAAATAGACACCTTAACATCAATTAAGAACTAGAGAAAGAG 4432
 Db 314 GlnGluArgSerLysIleAspThrLeuThrSerGlnLeuLysGluLeuGlnGlu 333
 QY 4433 CAACAAATCAAAAGCTAGCAGACACAAAGAAATAACTAAGATCAGACAGCAACTGAAG 4492
 Db 334 GlnThrHisSerLysAlaSerArgGlnGluIleThrLysIleArgAlaGluLeuLys 353
 QY 4493 GAGATAGACACACAAAAGCCCTTCAATTAATCAATCAATCAATCAATCAATCAATCAAT 4552
 Db 354 GluIleGluThrGlnLysThrLeuGlnLys-IleAsnGluSerArgSerTrpPheGlu 373
 QY 4553 AAGATCACAACAAAT---AGACCCTAGACAGACTAATAAGAAAGAAAGAGAGAA 4608
 Db 373 u***IleAsnLysIleAspArgProLeuAlaArgLeuIleLysLysLysArgLysLys 393
 QY 4609 TCAAGAGAGATCAATAAAAAATGATAAGGGGATATCACCCGATCCACAGCAATACA 4668
 Db 393 nGlnIleAspAlaIleLysAsnAspLysGlyAspIleThrThrAspProThrGluIleGlu 413
 QY 4669 AACTATTATCAGAGATATTATAAACACCTCTATGCAAAATAAACTAGAAAATCTAGAAGA 4728
 Db 413 nThrThrIleArgGluTyrTyrLysHisLeuTyrAlaAsnLysLeuGluAsnLeuGlu 433
 QY 4729 AATGGATAAATTCCTGGACACATATGTACCTGTATGGACCTTGGGGGAGACAGAAAG 4788
 Db 433 uMetAspLysPheLeuAsp----- 439
 QY 4789 GGGGTGAATGCAGAAATAAAAGACAAAGAGATGTTTGGAAAGTAGGGGTGAGG 4848
 Db 439 ----- 439
 QY 4849 GGGCAACTTGCCTTAATGGCAAGGGCCCTGAGCTTTACACCACCTCTGTATTATTA 4908
 Db 439 ----- 439
 QY 4909 GGCAAAAGAGATAGCGAGAGGTGAGTTGGAAAGAGAGGTGAGCTTTAGGTCCAGAGTA 4968
 Db 439 ----- 439
 QY 4969 GGCCTGCAAGACTGCATTCCTCAACAATAGGCTCTAGATGTCCTCCAGTAGATAACCTCAA 5028
 Db 439 ----- 439

QY 5029 GGAGCCAGTGCAGGGAGTGATGCCCTCAGCAAAACCTCTAGGGCAGGCACAGAAGTAA 5088
Db 439 ----- 439
QY 5089 GTTTGCCACACATTCTGTATTACGATAAACAGTTTGTGTTGTATCAAGTAGCCTCCAGT 5148
Db 439 ----- 439
QY 5149 GGAATGCTGAGTTGTCATGATCCCTTTGGCCCTTTTGGCTCCCAAAACACATACACCT 5208
Db 440 -----ThrTyrThrLe 443
QY 5209 CTCAGAGCTAAACACGGAAGAGTCAAAATCCCTGAATATACCAAGTAAAGTCTCAAAAT 5268
Db 443 uProArgLeuAsnGlnGluValGluSerLeuAsnArgProIleThrGlySerGluI 463
QY 5269 TGAAGCAGTAATTGATAGCCTTACCACCAACCAAAAAAGTCCAGGACCAAGGATTACAGC 5328
Db 463 e***AlaIleIleAsnSerLeuProThrLysLysSerProGlyProAspGlyPheThrAl 483
QY 5329 CAAATTTCTACAGAGGTACAAGAGAGCTGGTACTATTCTTCTGAAACTATTCCAAAA 5388
Db 483 aGluPheTyrGlnArgTyrLysGluGluValProPheLeuLeuLysLeuPheGlnSe 503
QY 5389 AATAGAA---AATGGAAATCCTCCTAACTACTATTTTACGAGGCCAGCATCTGTATAC 5445
Db 503 rIleGluLysGlu-GlyIleLeuProAsnSerPheTyrGluAlaSerIleIleLeuIleP 523
QY 5446 CAAACCTTAGCAGTGACACACAAAAAGAGGAATTTTCAGGCCCATCTCCTGTATGAACA 5505
Db 523 rLysProGlyArgAspThrThrLysLysGluAsnPheArgProIleSerLeuMetAsnI 543
QY 5506 TTGATGTGAAATCCTCAATAAAATACTGGCAAAACCAATCCAGCAGCATCAAAAAAGC 5565
Db 543 leAspAlaLysIleLeuAsnLysIleLeuAlaAsnArgIleGlnGlnHisIleLysLysL 563
QY 5566 TTATCTACCATGATCAAGTTGGCGTCATCCTCGGATGCAAGGCTGGTTCAAAATATGCA 5625
Db 563 euIleHisHisAspGlnValGlyPheIleProGlyMetGlnGlyTrpPheAsnIleArgL 583
QY 5626 AATCAATAAATGTAGGCCATCACATAAACAGAACCAATGTACAAAAACACATGATTATCT 5685
Db 583 ysSerIleAsnValIleGlnHisIleAsnArgThrLysAspLysAsnHisMetIleIleS 603
QY 5686 CAATAGATGCAAGAAAGCGCTTTCTCAAAATTCACAGCCCTTCATGCTCAAAATTTCTCA 5745
Db 603 erIleAspAlaGluLysAlaPheAspLysIleGlnGlnProPheMetLeuLysThrLeuA 623
QY 5746 GTAACTAGGTATCGATGGATGTATCTCAAAATTAAGAGCTATTATTATAC-AAACCCA 5804
Db 623 snLysLeuGlyIleAspGlyThrTyrLeuLysIleIleArgAlaIleTyrAspLysProt 643
QY 5805 CAGCAATATCATACTCAATGGGCAAAACTGGAAGCATTCCTCTTTCAGAACTGGCACAA 5864
Db 643 hrAlaAsnIleIleLeuAsnGlyGlnLysLeuGluAlaPheProLeuLysThrGlyThra 663
QY 5865 GACAAGGATGCCCTCTCTCACCCTCTATTCAGATACTATTGGAAAGTTCTGGCCAGGG 5924
Db 663 rgGlnGlyCysProLeuSerProLeuPheAsnIleValLeuGluValLeuAlaArgA 683
QY 5925 CAATCAGGCATAGAAAGATTAAGGGTATTCAATAGAAAGAGAGGAAGTATATTGT 5984
Db 683 laIleArgGlnGluLysGlyIleLysGlyIleGlnLeuGlyLysGluValLysLeuS 703
QY 5985 CTCTGTTGGCAGATGACATGTTGTATATTAGAAAAACCCATCGCTCTCAGGCCAAAAAC 6044
Db 703 erLeuPheAlaAspMetIleValTyrLeuGluAsnProIleValSerAlaGlnAsnL 723
QY 6045 TCCTTAAGCTGATAAGCAACTTCAGCAAAAGTCTCAGGACCAAAATCAATGTGCCAAAAT 6104
Db 723 euLeuLysLeuIleSerAsnPheSerLysValSerGlyTyrLysIleAsnValGlnLysS 743
QY 6105 CACAAGCATCTTATACGCCCAATAATAGACAAACAGAGAGCCCAATCATGAGTGAACCTCT 6164

Db 743 erGlnAlaPheLeuTyrThrAsnAsnArgGlnThrGluSerGlnIleMetSerGluLeuP 763
QY 6165 CATTCACAATTCGTACAAAGAGATAAAATACCTAGGAATACAACTTACAGGACACACGT 6224
Db 763 roPheThrIleAlaSerLysArgIleLysTyrLeuGlyIleGlnLeuThrArgaspValL 783
QY 6225 AGGAATCTTCAAGGAGAACTTACAACCACTGATCAAGGAAATTAAGAGAGACACAACA 6284
Db 783 ysAspLeuPheLysGluAsnTyrLysProLeuLeu***GluIleLysGluAspThrAsnL 803
QY 6285 AATGGAAAAACATTCCATGCTCAGATAGTAAGAATCAT-----GAAATGCCATAC 6337
Db 803 ysTrpLysAsnIleProCysSerTrpValGlyArgIleAsnIleValLysMetAlaIleL 823
QY 6338 TGCCCAAGATAAATTATAGATTACGTCTACCCCATCAAGCTACCACTGACTTTCCTTCA 6397
Db 823 euProLysValIleTyrArgPheAsnAlaIleProIleLysLeuProMetThrPhePhe 843
QY 6398 CAGAAATGGAAAAACAACTTTAAATTTTCATATGAACCAAAAAAGAGCCACAGAGCC 6457
Db 843 hrGluLeuGluLysThrThrLeuLysPheIleTrpAsnGln-LysArgAlaArgIleAla 862
QY 6458 AAGCAATCTTAAGCAAAAAGACAAAGCTGGAGGTATCATGCTACCTGACTTAAACTA 6517
Db 863 LysSerIleLeuSerGlnLysAsnLysAlaGlyIle***LeuProAspPheLysLeu 882
QY 6518 TACTATTAAGCTACAGTAACCAAACTGCATGGTACTGGTACCAAAAAACAGATATATAGC 6577
Db 883 TyrTyrLysAlaThrValThrLysThrAlaTrpTyrTrpTyrGlnAsnArgAspIleAsp 902
QY 6578 CAATGGAAACAGACAGAGCTCAGAAATTTACACT-GCAATCTACATCCCTGATCTT 6636
Db 903 GlnTrpAsnArgThrGluProSerGluIle***ProHisIleTyrAsn***LeuIlePhe 922
QY 6637 GACAACTCAGCAAAAACACAAAGCAATGGAAGAGATTCCCTTATTTAATAATGCTGTGG 6696
Db 923 AspLysPro***LysAsnLys***TrpGlyLysAspSerLeuPheAsnLysTrpCysTrp 942
QY 6697 AAAAATCGGTAGCATATGCAGAAAGCTGAAACTGGATCCCTTCCCTACACCTTATACA 6756
Db 943 GluAsnTrpLeuAlaIleCysArgLysLeuLysLeuAspProPheLeuThrProTyrThr 962
QY 6757 AAAGTTAACTCAAGATGAATTAAGACTTAAATATAAGACATAAAACCAATAAAACCCA 6815
Db 963 LysIleAsnSerArgTrpIleLysAspLeuAsnValArgProLysThrIleLysThrLeu 982
QY 6816 GAAGAAACCTTAGCCTATACCATTCAGGATATGCACATGGCAAGACTTTCATCACTAAA 6875
Db 983 GluGluAsnLeuGlyAsnThrIleGlnAspIleGlyMetGlyLysAspPheMet***Lys 1002
QY 6876 ACACCAAAAGCAATGGCAACAAAAGCCAAAATAGCAAGTGGGATCTGATTAAACTATAG 6935
Db 1003 ThrProLysAlaMetAlaThrLysAlaLysIleAspLysTrpAspLeuIleLysLeuLys 1022
QY 6936 AGCTTCGCACAGCAAAAAAACTGTCATCAGAGTGAAACAGCAACCTCAGAAATGGGA 6995
Db 1023 SerPheCysThrAla-LysGluThrThrIleArgValAsnArgGlnProThr***TrpGl 1042
QY 6996 GAAAATTTTGCATCTATCGATCTGACAAAGCTAATATCCAGAGATCTACGAAGAACT 7055
Db 1042 uLysIlePheAla***TyrSerAspLysGlyLeuIleSerArgIleTyr***GluLe 1062
QY 7056 TAAACAAATTTACAAGAAAAA---AACACCCCGCTCAAAATATGGCAAGGATATGAG 7111
Db 1062 uLysGlnIleTyrLysLysLysThrAsnAsnProIleLysLysTrpAlaLysAspMetAs 1082
QY 7112 CAGACACTTCTAAAAGACACATTTATGCAGGCCACAAACATATGAAAAAACCTCATC 7171
Db 1082 nArgHisPheSerLysGluAspIleTyrAlaAlaLys***HisMetLysLysCysSerSe 1102
QY 7172 ATCATTTGGTGTAGAGAAATGCAAAACAAACACAGTCACATACCATCTCATGCTAGT 7231
Db 7231 ATCATTTGGTGTAGAGAAATGCAAAACAAACACAGTCACATACCATCTCATGCTAGT 7231

Db 1102 rSerLeuAlaIleArgGluMetGlnIleLysThrThrMetArgTyThrHisLeuThrProVa 1122
QY 7232 TAGAATGGTGATCACCACAAAAGTCAGCAACAAATCTGCGAGAGGATGGAGAAAT 7291
Db 1122 lArgMetAlaIleIleLysSerGlyAsnAsnArgCysTrpArgGlyCysGlyGluI 1142
QY 7292 AGAACAATTTCCACTGTTGGTGGGAATGTAATTAAGTTCACCAATTTGGGAAGACAGT 7351
Db 1142 eGlyThrLeuLeuHisCysTrpAspCysLysLeuValGlnProLeuTriPylsSerVa 1162
QY 7352 GTGGAGATTCCTTAAGGATCTAGACCAAGAAATATCATTTGGCCAGCAATCCCATCT 7411
Db 1162 lTrpArgPheLeuArgAspLeuGluLeuGluIleProPheAspProAlaIleProLeu 1182
QY 7412 GAGTATATACCAAGGAATATAAATCATCTTATTATAAGACACATGCACACATATGTT 7471
Db 1182 uGlyIleTyProLysAspTyLysSerCysCysTyLysAspThrCysThrArgMetPh 1202
QY 7472 TATTGACAGCATGCATACAAAGACTTGGAAACCAACCAATGTCATCATCATGAT 7531
Db 1202 eIleAlaIleLeuPheThrIleAlaLysThrTrpAsnGlnProLysCysProThrMetI 1222
QY 7532 AGACTCGATAAGAAACATGGCACATATACACCATGAATACTATGCAGCCATAAAG 7591
Db 1222 eAspTrpIleLysLysMetTrpHisIleTyThrMetGluTyTyTrpAlaAlaIleLys 1242
QY 7592 -GATGATTCATGCTCTTTCAGAGATATGGATGAAGCTGGAAACCATCATCTTCAGCAA 7650
Db 1242 nAspGluPheMetSerPheValGlyThrTrpMetLysLeuGlu**IleIleLeuSerly 1262
QY 7651 ACTAACACAGAACAGAAACCAACACACATGTTCTCACTGTGAAGTGGAGT 7705
Db 1262 sLeuSerGln**GlnLysThrLysHisArg***PheSerLeuIleGlyGlyAsn 1280

RESULT 3

ABG06053 standard; Protein: 1284 AA.
XX AC ABG06053;
DT 13-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #6044.
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
PI WPI; 2001-639362/73.
DR N-PSDB; AAS70240.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX Claim 20; SEQ ID No 36412; 103pp; English.
PS The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 1284 AA;

Alignment Scores:

Pred. No.: 0 length: 1284
Score: 5164.00 Matches: 1049
Percent Similarity: 78.62% Conservative: 87
Best Local Similarity: 72.60% Mismatches: 143
Query Match: 9.87% Indels: 173
DB: 22 Gaps: 7

US-10-083-853B-2 (1-29921) x ABG06053 (1-1284)

QY 3401 CGGGCAAAATAACAGCTAGTGTCAATAATGGCAGGATCAAAATTCACACATAATAATTA 3460
Db 1 ArgAlaLysSerProAlaAsnIleIleMetThrGlySerAsnSerHisIleThrLeu 20
QY 3461 ACCTTAAATGTAATGGGTAAATTCCTCAATTAATAAGACACAGACTGCCAAATTTGGATA 3520
Db 21 ThrLeuAsnValAsnGlyLeuAsnSerProLysArgHisArgLeuAlaSerTrpIle 40
QY 3521 AAGATCAAGACCCATCAGTGTCTGTATTCAGGAGGCCATCTCATAGAAAGACACA 3580
Db 41 LysSerGlnAspProSerValCysCysIleGlnGluThrHisLeuMetCysArgAspThr 60
QY 3581 CATAGCTCAAAATAAAGGATGGAGAGATTTACCAAGTAAATCGAAACCAAAAAAAA 3640
Db 61 HisArgLeuLysIleLysGlyTrpArgLysIleTyTrpGlnAlaAsnGlyLysGln---Lys 79
QY 3641 AAAGCAGGGGTTGCAATCTCTAGTCTGATAAAACAGACTTTAAACCAACAAGATCAAA 3700
Db 80 LysAlaGlyValAlaIleLeuValSerAspLysThrAspPheLysProThrLysIleLys 99
QY 3701 AGAGCAAAAGAGGCCATTCATATATGTAAGGCATCAATGTAAGCAACAAGAGCTAAT 3760
Db 100 ArgAspLysGlyGlyHisTyMetMetValLysGlySerIleGlnGlnGluLeuThr 119
QY 3761 ATCCCTAAATATACATGCACCCCAATACAGGAGCCAGATTCATAGCAAGTCTCTAGA 3820
Db 120 IleLeuAsnMetTyAlaProAsnThrGlyAlaProArgPheIleLysGlnValLeuSer 139
QY 3821 GACCTACAAAGAGACTTTGACTCCCAACAATAATAGTGGAGTCTAAATAATAATAGA 3880
Db 140 AspLeuGlnArgAspLeuAspSerHisThrLeuIleMetGly-----Lys 153
QY 3881 CACTTTAACACCCCACTGCCAATATATAGCAGATCAATGAGACAGAGAAAAATTACAGGAT 3940
Db 154 AspPheAsnThrProLeuSerThrLeuAspArgSerThrArgGlnLysValAsnLysAsp 173
QY 3941 ATCCAGGAGTTGAAGTGAAGTCTGACCCAGCGGACCTAATAGATATCTACAGAACTCCC 4000
Db 174 ThrGlnGluLeuAsnSerAlaLeuHisGlnAlaAspLeuIleAspIleTyArgThrLeu 193

```
QY 4001 CACCCCAANTCAACAGATATATACACTCTCTCAGCATCATACATTAACCTATTTAAAT 4060
Db 194 HisProLysSerThrGluTyrThrPhePheSerAlaProHisHisSerTyrSerLysIle 213
QY 4061 GACCATGTAATTTAAAGTAAACACTCTCTCAGCAATCCAAAGCAAGAAATCCTAACA 4120
Db 214 AspHisIleLeuGlySerGluAlaLeuLeuSerLysCysLysArgThrGluIleIleThr 233
QY 4121 AACAGTCTCTCAGACTACAGTGCATCTATTAGAACTCAGAAATTAAGAAATCACTCAA 4180
Db 234 AsnTyrLeuSerAspHisSerAlaIleLysLeuGluLeuArgIleLysAsnLeuThrGln 253
QY 4181 AATCACACAACACTACATGAAGAACTGAACACCTCTCTCTGATGACTACTGGGTAAATAAC 4240
Db 254 SerArgSerThrTrpLysLeuAsnAsnLeuLeuLeuAsnAspTyrTrpValHisAsn 273
QY 4241 AAAATGAAGGCAAAATAAAGATCTCTTCAAAACCAATGAGACAAAGACACAAATGTAC 4300
Db 274 GluMetLysAlaGluIleLysMetPheGluThrAsnGluAsnLysAspThrThrTyr 293
QY 4301 CAGAACTCTCTGGGCATATTTAAAGCAGTGTGTAGAGGGAATTTATAGCAGCTAGATGCC 4360
Db 294 GlnAsnLeu**AspAlaPheLysAlaValcysargGlyLysPheIleAlaLeuAsnAla 313
QY 4361 TACAAGAGAAGCAGGAATATCTAAATAGACACCTTAACATCACAAATTAAGAACTA 4420
Db 314 HisLysArgLysGlnGluArgSerLysIleAspThrLeuThrSerGlnLeuLysGluLeu 333
QY 4421 GAGAAGAAAGCAACAAATTTCAAAAGCTAGCAGACAGACAGAAATAACTAAGATCAGA 4480
Db 334 GluLysGlnGluGlnThrHisSerLysAlaSerArgGlnGluIleThrLysIleArg 353
QY 4481 GCAGAAGTGAAGGAGATAGACACACAAAGCCCTTCAAAATAAATCAATGAATCCAGGAG 4540
Db 354 AlaGluLeuLysGluIleGluThrGlnLysThrLeuGlnLys-IleAsnGluSerArgSe 373
QY 4541 CTGGTTTTTCAAAAGATCAGCAAAAT----AGACCAGTACAGAGACTAATAAGAGAA 4596
Db 373 rTrpPheGluArgIleAsnLysIleAspArgProLeuAlaArgLeuIleLysLysLy 393
QY 4597 AAGAGAAGAAATCAAGAGATCAATATAAAATGATAAGGGGATATCACACCCGATCC 4656
Db 393 sArgGluLysAsnGlnIleAspThrIleLysAsnAspLysGlyAspIleThrThrAspPr 413
QY 4657 CACAGAATAACAACATTAATCAGAGAAATTAATAACACCTCTATGCAATAAACAATAGA 4716
Db 413 oThrGluIleGlnThrThrIleArgGluTyrTyrLysHisLeuTyrAlaAsnLysLeuGl 433
QY 4717 AAATCTAGAAGAAATGGATAAATTCCTGGACACATATGTAGCCCTGTATGGACCTTGGGG 4776
Db 433 uAsnLeuGluGluMetAspThrPheLeuAsp----- 443
QY 4777 ACAGAACAAAAGGGGTGAATGCAGAAATAAAGACAAAGACAAAGATGTTTGGAA 4836
Db 443 ----- 443
QY 4837 GTAGGGGTACGGGGCAACTTGCCTCTAATGACACAGGGCCCTGAGCTTTACACCCCT 4896
Db 443 ----- 443
QY 4897 CTGTATTTATTAGCAAAAGAGATAGCGAGAGGGTGAGTTGGAAGAAGAGGTACAGCTGT 4956
Db 443 ----- 443
QY 4957 AGGTCCAGAGTAGGCCTGCAAGACTGCATTCCTCAACAATAGGCTCTAGATGTCCAGT 5016
Db 443 ----- 443
QY 5017 AGATAACCTCAAGGAGCCAGTGCAGGGAGTGTATGGCCCTCAGCAAAACCTTCTAGGCGAG 5076
Db 443 ----- 443
QY 5077 GCACAGAAGTAAGTTTCCACATCTCTGATTACGATAAACAGTTTGGCTTGTATGATCAA 5136
```

```
Db 443 ----- 443
QY 5137 GTAGCCTCCAGTGAATGCTGAGTTGTCATGATCCCTTTGGCCCTTTTGGCTCCCAAAA 5196
Db 443 ----- 443
QY 5197 CACATACACCCCTCTCAAGACTAAACACAGGAAGTCAAAATCCCTGAATATACAGTAAC 5256
Db 444 -ThrTyrThrLeuProArgLeuAsnGlnGluValGluSerLeuAsnArgProIleTh 463
QY 5257 AAGTTCTAAATTCGAACGAGTAATTGATAGCCTACCACCAACCAAAAAGTCCAGGACACAGA 5316
Db 463 rGlyAlaGluIleValAlaIleIleAsnSerLeuSerThrLysLysSerProGlyProAs 483
QY 5317 CGGATTCCACAGCCCAAAATCTACCAGAGGTACAAAGAGAAGCTGTACTATTCTCTCGAA 5376
Db 483 pGlyPheThrAlaGluPheTyrGlnArgTyrLysGluGluLeuValProPhePheLeuLy 503
QY 5377 ACTATTCACAAAAAATAGAA---AATGGAAATCCCTCCCTAACTATTTCAGAGGCCACGA 5433
Db 503 sLeuPheGlnSerIleGluLysGlu-GlyIleLeuProAsnSerPheTyrGluAlaSerI 523
QY 5434 TCATCTGTATACCAAAACCTAGCAGTGCACACAAACAAAGAGGAAATTCAGGCCCATAT 5493
Db 523 leIleLeuIleProLysProGlyArgAspThrThrLysLysGluAsnPheArgProIleS 543
QY 5494 CCCTGATGAACATTGTGAATAATCCTCAATAAATACTGGCAAAACCAATCCAGCAGC 5553
Db 543 erLeuMetAsnIleAspAlaLysIleLeuAsnLysIleLeuAlaAsnGlnIleGlnGlnH 563
QY 5554 ACATCAAAAAGCTTATCTACCATGATCAAGTGGCGTCATCCCTGGGATGCAAGCGTGGT 5613
Db 563 isIleLysLysLeuIleHisAspGlnValGlyPheIleProGlyMetGlnGlyTrp 583
QY 5614 TCANAATATGCAATCAATAAATGTAGCCCATCATATAACAGAACCAATGACAAAACACC 5673
Db 583 heAsnIleArgLysSerIleAsnValIleGlnHisIleAsnArgAlaLysAspLysAsnH 603
QY 5674 ACATGATATTCTCAATAGATGCAGAAAGCGCTTTGTCAAAATTTCAACAGCCCTTCATGC 5733
Db 603 sMetIleValSerIleAspAlaGluLysAlaPheAspLysIleGlnGlnProPheMetL 623
QY 5734 TAAAAATCTCAGTAACACTAGGTATCGATGGAATGTATCTCAAAATAATTAAGACTATTT 5793
Db 623 euLysThrIleAsnLysIleGlyIleAspGlyMetTyrPheLysIleIleArgAlaIleT 643
QY 5794 ATAC-AAACCCACAGCAATATCATCTGAATGGGCAAAACTGCAAGCATTCCTCTTTGA 5852
Db 643 yrAspLysThrThrAlaAsnIleIleLeuAsnGlyGlnLysLeuGluAlaPheProLeuL 663
QY 5853 GAATGGCCACAGCAAGGATGCCCTCTCTCACCACCTCCTATTCAGATATACTATTGGAAG 5912
Db 663 ystThrGlyThrArgGlnGlyCysProLeuSerProLeuLeuPheAsnIleValLeuGluP 683
QY 5913 TTCTGGCCAGGGCAATCAGCAATAGAAAGAAATAAAGGGTATTCAATAGAAACAGACAGG 5972
Db 683 heLeuAlaArgAlaIleArgGlnGluLysGluIleLysGlyIleGlnLeuGlyLysGluG 703
QY 5973 AAGTCATATTGCTCTCTCTTTGGCAGATGACATGTTTGTATATTTTGAAGAACCCCATCTCT 6032
Db 703 luValLysLeuSerLeuPheAlaAspMetIleValAspLeuGluAsnProIleValS 723
QY 6033 CAGGCCAAAACCTCTTAAGCTGATAAGCAACTTCAGCAAAAGTCTCAGGACACAAAATCA 6092
Db 723 erAlaGlnAsnLeuLeuLysLeuIleSerAsnPheSerLysLeuSerGlyTyrLysIleA 743
QY 6093 ATGTGCAAAATCACAAGCATTTCTTATCGCCCAATAATAGACAAACAGAGACCAATCA 6152
Db 743 snValGlnLysSerGlnAlaPheLeuTyrThrAsnAsnArgGlnThrGluSerGlnIleM 763
QY 6153 TGAGTGAACCTCTCATCACAATTCCTACAAAGAGAATAAATACCTAGGATACAACTTA 6212
Db -----
```

Db 763 etSerGluLeuProPheThrIleAlaSerLysArgIleLysTyrLeuGlyIleGlnLeu 783
 QY 6213 CAAGGACACCTAGGAACTCTCAAGGAGAACTACAAACACCTGATCAAGGAAATAGAG 6272
 Db 783 hrArgspValLysAspLeuPheLysGluAsnTyrLysProLeuLeuLysGluIleLysG 803
 QY 6273 AGGACAAACAATAGGAAACATTCCTACGATCAGATAGTAAGTAATCAT-----G 6325
 Db 803 luAspThrLysTyrLysAsnThrProCysSerTrpValGlyArgIleAsnIleVal 823
 QY 6326 AAATGCCATACGCCCAAGTAATATATAGATTCAGTCTACCCCTCAAGCTACCAT 6385
 Db 823 ysMetAlaIleLeuProLysValIleTyrArgPheAsnAlaIleProIleLysLeuProm 843
 QY 6386 TGACTTTCTTCACAGAAATCGAAAAACAACATTTAAATTTCAATATGGAACCAAAAAAGA 6445
 Db 843 etThrPheThrGluLeuGluLysThrThrLeuLysPheIleThrPasnGln-LysArg 862
 QY 6446 GCCCACAGCCCAAGCAATCTTAAGCAAAAAGAACAAAGCTGGAGTATCATGCTACCT 6505
 Db 863 AlaArgIleAlaLysSerIleLeuSerGlnLysAsnLysAlaGlyGlyIleThrLeuPro 882
 QY 6506 GACTTAAACTATATAGCTACAGTAACCAAACTCGATGGTACTGCTACCAAAAC 6565
 Db 883 AspPheLysLeuTyrLysAlaThrValThrLysThrAlaTyrTyrTrpTyrGlnAsn 902
 QY 6566 AGATATATAGCAATGGAACAGACAGACCTCAGAAATTTACACT-GCAATCTACATC 6624
 Db 903 ArgAspIleAspGlnTrpAsnArgThrGluProSerGluIleThrProLeuThrTyrAsn 922
 QY 6625 CATCTGATCTTGCACAACTGACAAAACAAAGCAATGGAAGAGTTCCTTATTAAT 6684
 Db 923 TyrLeuIlePheAspLysProGluLysAsnLysGlnTrpGlyLysAspSerLeuPheAsn 942
 QY 6685 AAATGGTGTGGAAAACCTGGCTAGCCATATGAGCAAGCTGAACTGGATCCCTCCTT 6744
 Db 943 LysTrpCysTrpGluAsnTrpLeuAlaIleCysArgLysLeuLysValAspProPheLeu 962
 QY 6745 ACACCTTATACAAAGTTAACTCAAGATGAATTAAGACTTAAATATAAGACATAAAACC 6804
 Db 963 ThrProTyrThrLysIleAsnSerArgTrpIleLysAspLeuAsnValArgProLysThr 982
 QY 6805 ATAAACCCCA-GAAGAAAACCTAGGCAATACCATTACGATATGGACATGGGCAAGAC 6863
 Db 983 IleLysThrLeuGluGluAsnLeuGlyIleThrIleGlnAspIleGlyValGlyLysAsp 1002
 QY 6864 TTCATGACTAAACACCAAGCAATGGCAACAAACCCCAAAATAGACAGTGGGATCTG 6923
 Db 1003 PheMetSerLysThrProLysAlaMetAlaThrLysAlaLysIleAspLysTrpAspLeu 1022
 QY 6924 ATAAACTATAGACTTCTGCACAGCAAAAAAACTGTCTATCAGAGTGAACAGCAACC 6983
 Db 1023 IleLysLeuLysSerPheCysThrAla-LysGluThrThrIleArgValAsnArgGlnPr 1042
 QY 6984 TACAGATGGAGAAAATTTTGCATCTATCGATCTGCAAAAGCTAATATATCCAGAGAT 7043
 Db 1042 oThrThrTrpGluThrIlePheThrThrTyrSerSerAspLysGlyLeuIleSerArgII 1062
 QY 7044 CTACGAGAACTTAACAAATTTTACAGAAAAA- - -AACAAACCCCTCAAAATATGGGC 7099
 Db 1062 eTyrAsnGluLeuLysGlnIleTyrLysLysLysThrAsnAsnProIleLysLysTrpAl 1082
 QY 7100 AAAGGATATGAGCAGACACTTCTCAAAAGAAAGACATTTATGACGCAACCAACATATGAA 7159
 Db 1082 aLysAspMetAsnArgHisPheSerLysGluAspIleTyrAlaAlaLysLysHisMetLy 1102
 QY 7160 AAAAACCTCATCATCATCTGCTTGTAGAGAAATTCAAAACAAACCAACAGCTGACATACCA 7219
 Db 1102 sLysCysSerSerSerLeuAlaIleArgGluMetGlnIleLysThrThrMetArgTyrHI 1122
 QY 7220 TCTCATGCTAGTATGATGCTGATCACTAAAGTACGAGAAACAAACAAATCTGGAGAGG 7279
 Db 1122 sLeuThrProValArgMetAlaIleIleLysLysSerGlyAsnAsnArgCysTrpArgGI 1142

QY 7280 ATGTGAGAAATAGGAACACTTTTCCACTGTGTGGGAATGTAAATTTAGTTCAACCAT 7339
 Db 1142 yCysGlyGluIleGlyThrLeuLeuHisCysTyrTrpAspCysLysLeuValGlnProLe 1162
 QY 7340 GTGAGACAGCTGGAGATTCCTTAAGATCTAGAACCAAGAAATATCATTTGACCAGC 7399
 Db 1162 uTrpLysSerValTrpArgPheLeuArgAspLeuGluLeuGluIleProPheAspProAl 1182
 QY 7400 AATCCCATCTACTGAGTATATACCAAGGAATATAATCATCTTATATATAAGACACATG 7459
 Db 1182 aIleProLeuLeuGlyIleTyrProAsnAspTyrLysSerCysCysTyrLysAspThrCy 1202
 QY 7460 CACACATATGTTTATTCGACACTGATCACAATAGCAAAAGACTTGAACCAACCAATG 7519
 Db 1202 sThrArgMetPheIleAlaLeuPheThrIleAlaLysThrTrpAsnGlnProLysCy 1222
 QY 7520 TCATCAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 7579
 Db 1222 sProThrMetIleAspTrpIleLysLysMetTrpHisIleTyrThrMetGluTyrTrAl 1242
 QY 7580 AGCCATAAAAG-CATGAGTTCATGCTCTTTCAGAGATATGATGATGATGATGATG 7638
 Db 1242 aAlaIleLysAsnAspGluPheMetSerPheValGlyThrTrpMetLysLeuGluThrII 1262
 QY 7639 CATTCAGCAAACTAACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 7698
 Db 1262 eIleLeuSerLysLeuSerGlnGluLysThrLysHisArgIlePheSerLeuIleGI 1282
 QY 7699 TGGGAGT 7705
 Db 1282 yGlyAsn 1284

RESULT 4
 ABG09636
 ID ABG09636 standard; Protein; 1284 AA.
 AC ABG09636;
 XX
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #9627.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN W0200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US086631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS73823.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 XX
 PS Claim 20; SEQ ID NO 39995; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX
 SQ Sequence 1284 AA;

Alignment Scores:

Pred. No.:	0	Length:	1284
Score:	5164.00	Matches:	1049
Percent Similarity:	78.62%	Conservative:	87
Best Local Similarity:	72.60%	Mismatches:	143
Query Match:	9.87%	Indels:	173
DB:	22	Gaps:	7

US-10-083-853B-2 (1-29921) x ABG09636 (1-1284)

QY	3401	CGGCAAAATACACGCTAGTGTCTAATGGCAGGATCAAAATTCACATATAATATTA	3460
DB	1	ArgAlaLysSerProAlaAsnIleMetThrGlySerAsnSerHisIleThrIleLeu	20
QY	3461	ACCTTAATGTAAATGGCTAAATTCGCCAATAAAGACACAGACTGGCAATTTGGATA	3520
DB	21	ThrLeuAsnValAsnGlyLeuAsnSerProIleTyrArgHisArgLeuAlaSerTrpIle	40
QY	3521	AAGACTCAAGCCCATCAGTGTCTGTATTCAGGAGGCCATCTCACATGAAGAACACACA	3580
DB	41	LysSerGlnAspProSerValCysIleGlnGluThrHisLeuMetCysArgAspThr	60
QY	3581	CATAGGCTCAAAATAAAGGATGGAGGAGATTTACCAAGTAATGGAAACAAACAAAA	3640
DB	61	HisArgLeuLysIleLysGlyTrpArgLysIleTyrGlnAlaAsnGlyLysGln--Lys	79
QY	3641	AAAGCAGGGTTGCATCCTAGTCTCTGATAAACAGACTTTAAACCAACAAAGATCAA	3700
DB	80	LysAlaGlyValAlaIleLeuValSerAspLysThrAspPheLysProThrLysIleLys	99
QY	3701	AGAGCAAGAGGCCATTACATATGTAAAGGCATCAATGGACAGAGAGAGCTAACT	3760
DB	100	ArgAspLysGluGlyHisTyrMetMetValLysGlySerIleGlnGlnGluLeuThr	119
QY	3761	ATCCTAAATATACATGCCCAATACAGAGCACCCAGATTCATAAGCAAGTCTTTAGA	3820
DB	120	IleLeuAsnMetTyrAlaProAsnThrGlyAlaProArgPheIleLysGlnValLeuSer	139
QY	3821	GACCTACAAAGAGACTTTGACTCCACACATAATAGTGGGAGTCTAAATAATAATAGA	3880
DB	140	AspLeuGlnArgAspLeuAspSerHisThrIleMetGly-----	153
QY	3881	CACTTTAAACCCCACTGCCAATATTAGGCGAGATCAATGAGACAGAAAAATTAACAAGAT	3940
DB	154	AspPheAsnThrProLeuSerThrLeuAspArgSerThrArgGlnLysValAsnLysAsp	173
QY	3941	ATCCAGGAGTTGAAGTACGCTGGACCAAGCGGACCTTAATAGATATCTACAGAGATCCC	4000
DB	174	ThrGlnGluLeuAsnSerAlaLeuHisGlnAlaAspLeuIleAspIleTyrArgThrLeu	193
QY	4001	CACCCCAATCAACAGATATACACTCTTCTCAGCATCACATTTACACCTATTTTAAATTT	4060

DB	194	HisProLysSerThrGluTyrThrPhePheSerAlaProHisHisSerTyrSerLysIle	213
QY	4061	GACCATGTAAATTTAAGTAAACACTCCTCAGCAAAATGCAAAAGAACAAATCTCAACA	4120
DB	214	AspHisIleLeuGlySerGluAlaLeuLeuSerLysCysLysArgThrGluIleIleThr	233
QY	4121	AACAGTCTCTCAGACTACAGTGCATATTTAGAACTCAGAAATTAAGAACTACTCAAA	4180
DB	234	AsnTyrLeuSerAspHisSerAlaIleLysLeuGluLeuArgIleLysAsnLeuThrGln	253
QY	4181	AATCACAACTACATGGAACCTGCTCCTCGAATGACTACTGGGTAATAATAC	4240
DB	254	SerArgSerThrThrTrpLysLeuAsnAsnLeuLeuLeuLeuAsnAspTyrTrpValHis	273
QY	4241	AAATGAAGGCAAAATAAAGATGTCTTGAACCAATGAGCAACAAAGACACATGTAC	4300
DB	274	GluMetLysAlaGluIleLysMetPheGluThrAsnGluAsnGluAspThrThrTyr	293
QY	4301	CAGAACTCTGGGGCATATTTAAAGCAGTGTAGAGGGAATTTATAGCACTAGATGCC	4360
DB	294	GlnAsnLeu***AspAlaPheLysAlaValCysArgGlyLysPheIleAlaLeuAsnAla	313
QY	4361	TACAAGAGAAAGCAGGAATATCTAAATAGACACCTTAACATCAATTAAGAACTA	4420
DB	314	HisLysArgLysGlnGluArgSerLysIleAspThrLeuThrSerGlnLeuLysGluLeu	333
QY	4421	GAGAAGAAAGACGCAAAATTTCAAAAGCTAGCAGACAGCAAGAAATTAAGATCAGA	4480
DB	334	GluLysGlnGluGlnThrHisSerLysAlaSerArgArgGlnGluIleThrLysIleArg	353
QY	4481	GCAGAACTGAAGGATAGACACACAAAAGCCCTCAAAATTAATCAATGAATCAGAG	4540
DB	354	AlaGluLeuLysGluIleGluThrGlnLysThrLeuGlnLys--IleAsnGluSerArgSe	373
QY	4541	CTGGTTTTTCAAAAGATCAGCAAAAT-----AGACCCTAGACACTAAATAAGAA	4596
DB	373	rTTPheGluArgIleAsnLysIleAspArgProLeuAlaArgLeuIleLysLysLys	393
QY	4597	AAGAGAGAAAGTCAAGAGATGCAATAAAAATGATAAAGGGATATCACCACCGATCC	4656
DB	393	sArgGluLysAsnGlnIleAspThrIleLysAsnAspLysGlyAspIleThrAspPr	413
QY	4657	CACAGAAATCAAACTATTATCAGAGAAATATTATAACACCTCTATGCAATAAATACTAGA	4716
DB	413	oThrGluIleGlnThrThrIleArgGluTyrTyrLysHisLeuTyrAlaAsnLysLeuG	433
QY	4717	AAATCTAGAAATGGATAAATTCCTGGACACATATGTAGCCTGTATGGACCTTGGGGG	4776
DB	433	uAsnLeuGluGluMetAspThrPheLeuAsp-----	443
QY	4777	ACAGAACAAAAGGGGTGAATGCAGAAATAAAGACAAAGAGATGTGTTGAA	4836
DB	443	-----	443
QY	4837	GTAGGGGTGAGGGGCAACTTGCTCTAATGACAAAGGGCCCTGAGCTTTACACCCCT	4896
DB	443	-----	443
QY	4897	CTGTATTATTATAGCAAAAGAGATAGCCAGAGGGTGAGTTGGAAGAAGAGGTCAGCTGT	4956
DB	443	-----	443
QY	4957	AGGTCCAGAGTAGGCTGCAAGACTGCATTCCTCAACAATAGGCTCTAGATGTCCCACT	5016
DB	443	-----	443
QY	5017	AGATAACCTCAAGGAGCGAGTGCCAGGGAGTGATGGCCCTCAGCAAACTTCTTAGGGCAG	5076
DB	443	-----	443
QY	5077	GCACAGAAGTAAGTTTGCCCACTTCTGTATTATTCAGATAAACAGTTTCTGCTTTGATCAA	5136

Db	443	-----	443
Qy	5137	GTAGCCTCCAGTGAATGCTGAGTGTGTCATGCTCCTTTGGCCTTTGGCTCCCAAAA	5196
Db	443	-----	443
Qy	5197	CACATACACCTCTCAAGACTAAACACGAGAGTCAAAATCCCTGTAATATACACGTAAC	5256
Db	444	-ThrTyr-ThrLeuProArgLeuAsnGlnGluValGluSerLeuAsnArgProIleTh	463
Qy	5257	AAGTTCCTAAATCAACGACGTAATGTATAGCCTACCAACCAAAAAAGTCCAGGACAGA	5316
Db	463	IGlyAlaGluValAlaIleLeuSerLeuThrLysLysSerProGlyProAs	483
Qy	5317	CGGATTCACAGCCAAATCTTACACGAGGTCAAGAGAACTGTACTATTCTCTCTGAA	5376
Db	483	PGlyPheThrAlaGluPheTyrGlnArgTyrLysGluGluLeuValProPhePheLeuLy	503
Qy	5377	ACTATTCCTCAAAAATAGAA---AATGGGAATCTCTCCCTAACTCATTTTACGAGCCAGCA	5433
Db	503	sLeuPheGlnSerIleGluLysGlu-GlyIleLeuProAsnSerPheTyrGluAlaSerI	523
Qy	5434	TCATCCTGATACCAAAACCTAGCAGTCACACACAAACAAAGAGGAAATTTTCAGGCCATAT	5493
Db	523	IleIleLeuIleProLysProGlyArgAspThrThrLysLysGluAsnPheArgProIleS	543
Qy	5494	CCCTGATGAACATGTGATGCTGAAATCCTCAATAAATPACTGGCAAAACCAATCCAGCAGC	5553
Db	543	erLeuMetAsnIleAspAlaLysIleLeuAsnLysIleLeuAlaAsnGlnIleGlnGlnH	563
Qy	5554	ACATCAAAAAGCTTATCTACCATGATCAAGTGGCGTCATCCCTGGGATCCAAAGCTGGT	5613
Db	563	isIleLysLysLeuIleHisAspGlnValGlyPheIleProGlyMetGlnGlyTrp	583
Qy	5614	TCRAAATATCAATCAATCAATCTAGCCCATCACATAACAGACCAACCAATGCACAAAACC	5673
Db	583	heAsnIleArgLysSerIleAsnValIleGlnHisIleAsnArgAlaLysAspLysAsnH	603
Qy	5674	ACATGATATCTCAATAGATGCAGAAAGCCCTTTGTCAAAATCAACAGCCCTTCATGC	5733
Db	603	isMetIleValSerIleAspAlaGluLysAlaPheAspLysIleGlnGlnProPheMetL	623
Qy	5734	TAAAAATTCAGTAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	5793
Db	623	eulysThrIleAsnLysIleGlyIleAspGlyMetTyrPheLysIleIleArgAlaIleT	643
Qy	5794	ATAC-AAACCCACAGCAATATCATCTGAATGGCAAAACCTGGAAGCATTCCTTTGA	5852
Db	643	yrAspLysThrThrAlaAsnIleIleLeuAsnGlyGlnLysLeuGluAlaPheProLeuL	663
Qy	5853	GAAGTGGCACAAGACAGGATGCCCTCTCTCACCATCTCTTATCAAGATACTATTGGAAG	5912
Db	663	ysThrGlyThrArgGlnGlyCysProLeuSerProLeuLeuPheAsnIleValLeuGluP	683
Qy	5913	TTCTGGCCAGGCAATCAGGCATAGAAAGAAATAAAGGTATTCAAATAGAAAGAGG	5972
Db	683	heLeuAlaArgAlaIleArgGlnGluLysGlyIleGlnLeuGlyLysGluG	703
Qy	5973	AAGTCATATGTCTCTCTTTGCAGATGACATGTTGTGTATTTAGAAACCCCATCTCTCT	6032
Db	703	IuValLysLeuSerLeuPheAlaAspMetIleValAspLeuGluAsnProIleValS	723
Qy	6033	CAGGCCAAAACCTCTTAAGCTGATAGCACTTCAGCAAGTCTCAGGACACAAAATCA	6092
Db	723	erAlaGlnAsnLeuLeuLysLeuIleSerAsnPheSerLysLeuSerGlyTyrLysIleA	743
Qy	6093	ATGTGCAAAATCACAAGCATTTCTTATACGCCAATATAGACAAACAGAGAGCCAAATCA	6152
Db	743	snValGlnLysSerGlnAlaPheLeuTyrThrAsnAsnArgGlnThrGluSerGlnIleM	763
Qy	6153	TGAGTGAACCTCTATTACAAATTCCTACAAGAGATAAATACCTAGGAATACAACTTA	6212
Db	763	etSerGluLeuProPheThrIleAlaSerLysArgIleLysTyrLeuGlyIleGlnLeuT	783
Qy	6213	CAAGGGACACGTAGAACTCTTCAAGGAGAACTACAAACCACTGATCAAGAAATAAGAG	6272
Db	783	hrArgspValLysAspLeuPheLysGluAsnTyrLysProLeuLeuLysGluIleLysG	803
Qy	6273	AGGACACAAACAATGAAACCAATTCATGCTCACAGATAGTAGAATCAT-----G	6325
Db	803	IuAspThrLysLysTrpLysAsnThrProCysSerTrpValGlyArgIleAsnIleValL	823
Qy	6326	AAATGCCATCTACCCCAAGTAATATAGATTCAGTGTACCTACCCCATCAAGTACCAT	6385
Db	823	ysMetAlaIleLeuProLysValIleTyrArgPheAsnAlaIleProIleLysLeuProM	843
Qy	6386	TGACTTTCTTCACAGAAATTCGAAAAACAACTTTTAAATTTTCATATGCAACCAAAAAAGA	6445
Db	843	etThrPheThrGluLeuGluLysThrThrLeuLysPheIleTrpAsnGln-LysArg	862
Qy	6446	GCCACAGACCCAGACAATCTTAAGCAAAAGAACAAAGCTGGAGGTATCATCTACCT	6505
Db	863	AlaArgIleAlaLysSerIleLeuSerGlnLysAsnLysAlaGlyGlyIleThrLeuPro	882
Qy	6506	GACTTAAACATATATAGGCTACAGTACCAACCAACTGCATGTACTGTACCAAAAC	6565
Db	883	AspPheLysLeuTyrTyrLysAlaIleThrValThrLysThrAlaTrpTyrTrpTyrGlnAsn	902
Qy	6566	AGATATATAGACCAATGGAACAGAACAGACCTCAGAAATTTACACT-GCAATCTACATC	6624
Db	903	ArgAspIleAspGlnTrpAsnArgThrGluProSerGluIleThrProLeuThrTyrAsn	922
Qy	6625	CATCTGATCTTGCACAACTGCACAAACAGCAATGGAAGAAAGATTCCTATTAAAT	6684
Db	923	TyrLeuIlePheAspLysProGluLysAsnLysGlnTrpGlyLysAspSerLeuPheAsn	942
Qy	6685	AAATGGTGTGGAAAACTGCTAGCCATATGCAAGAACTGAAACCTGGATCCCTCTCTT	6744
Db	943	LysTrpCysTrpGluAsnTrpLeuAlaIleCysArgLysLeuLysValAspProPheLeu	962
Qy	6745	ACACCTTATACAAAGTTAACTCAAGATGAATTAAGACTTAATAATAAGACATAAAACC	6804
Db	963	ThrProTyrThrLysIleAsnSerArgTrpIleLysAspLeuAsnValArgProLysThr	982
Qy	6805	ATAAAACCCA-GAAGAAACCTAGGCAATACCATTCAGGATATGGACATGGGCAACAC	6863
Db	983	IleLysThrLeuGluGluAsnLeuGlyIleThrIleGlnAspIleGlyValGlyLysAsp	1002
Qy	6864	TTCATGACTAAAACACCAAGCAATGGCAACAAACCCCAAAATAGACAAGTGGATCTG	6923
Db	1003	PheMetSerLysThrProLysAlaMetAlaThrLysAlaLysIleAspLysTrpAspLeu	1022
Qy	6924	ATTAAACTATAGACTTCTGCACAGCAAAAAAACTGTCTCATCAGAGTGAACAGCAACC	6983
Db	1023	IleLysLeuLysSerPheCysThrAla-LysGluThrThrIleArgValAsnArgGlnPr	1042
Qy	6984	TACAGAATGGAGAAAATTTTGCATCTCATGATCTGCAAAAGGCTTAATATCCAGAGAT	7043
Db	1042	oThrThrTrpGluThrIlePheThrThrTyrSerSerAspLysGlyLeuIleSerArgIl	1062
Qy	7044	CTACGAAGAATTAACAAAATTTTACAGAAAAA- - -AACACCCCTCAAAATATCGGC	7099
Db	1062	eTyrAsnGluLeuLysGlnIleTyrLysLysLysThrAsnAsnProIleLysLysTrpAl	1082
Qy	7100	AAAGGATATAGGACACACTTCTCAAAAGAAAGACATTTATGCAACCAACCAATATCAA	7159
Db	1082	aLysAspMetAsnArgHisPheSerLysGluAspIleTyrAlaAlaLysLysHisMetLy	1102
Qy	7160	AAAAACCTCATCATCATTTGCTGTTAGAGAAATGCAAAACAAACCAACACACATACCA	7219
Db	1102	sLysCysSerSerSerLeuAlaIleArgGluMetGlnIleLysThrThrMetArgTyrHi	1122
Qy	7220	TCATCATCTAGTGTAGATGCTGATCACTAAAGATCAGGAAACAAACAAATCTGGAGAG	7279
Db	1122	sLeuThrProValArgMetAlaIleLysLysSerGlyAsnAsnArgCysTrpArgGl	1142


```
QY 7280 ATGTGGAGAAATAGGAACACTTTCCACTGTTGGGGAATATAATAGTTCACCACTT 7339
Db 1142 yCyseGlyGluIleGlyThrLeuLeuHisCysTrpAspCysLysLeuValGlnProle 1162
QY 7340 GTGGAAGACAGTGTGGAGATTCCTTAAGGATCTAGAACACAGAAATATCATTTGACCCAGC 7399
Db 1162 utrplysservaltPArgPheLeuArgAspLeuGluLeuGluIleProPheAspProAl 1182
QY 7400 AATCCATTACTAGTATATACCCAAAGGAATATAAATCATCTCTATTATAAGACACATG 7459
Db 1182 alleProLeuLeuGlyIleTyProAsnAspTrpLysSerCysTyTrpLysAspThCy 1202
QY 7460 CACACATATGTTTATGAGCACTGATCACAATAGCAAGACTTGGACCAACCAATG 7519
Db 1202 sThrArgMetPheIleAlaLeuPheThrIleAlaLysThrTrpAsnGlnProLysCy 1222
QY 7520 TCATCAGTGTAGACTGGATAAGAAACATGGGCACATATACACCATGAATACATATGC 7579
Db 1222 sProThrMetIleAspTrpIleLysLysMetTrpHisIleTyThrMetGluTyTrAl 1242
QY 7580 AGCCATAAAAG-GATGAGTTTCATGCTCTTTTCAGAGATATGGATGAAGCTGGAACCAT 7638
Db 1242 aAlaIleLysAsnAspGluPheMetSerPheValGlyThrTrpMetLysLeuGluThrIl 1262
QY 7639 CATTCAGCAACTAATACACAGAACAGAAACCAACCAACCACTGTTCTCACTGTGTAAG 7698
Db 1262 eIleLeuSerLysLeuSerGlnGluGlnLysThrLysHisArgIlePheSerLeuIleGl 1282
QY 7699 TGGGAGT 7705
Db 1282 yGlyAsn 1284

RESULT 5
ABG10795
ID ABG10795 standard; Protein; 1284 AA.
AC ABG10795;
XX
XX
DT 13-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #10786.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX PD 11-OCT-2001.
XX
XX PF 30-MAR-2001; 2001WO-US08631.
XX
XX PR 31-MAR-2000; 2000US-0540217.
XX
XX PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX
XX DR N-FSDB; AAS74982.
```

New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.

Claim 20; SEQ ID No 41154; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1284 AA;

Alignment Scores:
Pred. No.: 0 Length: 1284
Score: 5164.00 Matches: 1049
Percent Similarity: 78.62% Conservative: 87
Best Local Similarity: 72.60% Mismatches: 143
Query Match: 9.87% Indels: 173
DB: 22 Gaps: 7

US-10-083-853b-2 (1-29921) x ABG10795 (1-1284)

```
QY 3401 CGGCAAAATAACACGCTAGTGTCTATAATGGCAGGATCAAAATTCACATATAATATTA 3460
Db 1 ArgAlaLysSerProAlaAsnIleIleMetThrGlySerAsnSerHisIleThrLeu 20
QY 3461 ACCTTAATGTAAATGGCTAAATTCCTCCCAATTAAGACACACAGCTGGCAATTTGGATA 3520
Db 21 ThrLeuAsnValAsnGlyLeuAsnSerProIleLysArgHisArgLeuAlaSerTrpIle 40
QY 3521 AAGAGTCAAGACCCATCAGTGTCTGTATTTCAGGAGGCCCTCTCACATGAACACACACA 3580
Db 41 LysSerGlnAspProSerValCysIleGlnGluThrHisLeuMetCysArgAspThr 60
QY 3581 CATAGGCTCAAAATAAAGGGATGGAGGAAGATTTACCAAGTAATTTGGAACCAACAAAAA 3640
Db 61 HisArgLeuLysIleLysGlyTrpArgLysIleTyArgGlnAlaAsnGlyLysGln---Lys 79
QY 3641 AAAGCAGGGGTGCAATCCTAGTCTCTGATATAAACACAGACTTTAAACCAACAAAGATCAA 3700
Db 80 LysAlaGlyValAlaIleLeuValSerAspLysThrAspPheLysProThrLysIleLys 99
QY 3701 AGAGCAAAAGAGGCGCATTACATTAATGCTAAAGGCATCAATGGCAACAAGAGAGTAACT 3760
Db 100 ArgAspLysGluGlyHisTyMetMetValLysGlySerIleGlnGlnGluLeuThr 119
QY 3761 ATCCTAAATATACATGCACCCCAATACAGGAGCACCAGATTTCATAAAGCAAGTCTTTAGA 3820
Db 120 IleLeuAsnMetTyAlaProAsnThrGlyAlaProArgPheIleLysGlnValLeuSer 139
QY 3821 GACCTACAAGAGACTTTGACTCCACACATATATAGTGGGAGTCTAAATATAATAATAGA 3880
Db 140 AspLeuGlnArgAspLeuAspSerHisThrLeuIleMetGly----- 153
QY 3881 CACTTTTAACACCCCACTGCCAATATTAGGCGCATCAATAGCAAGAGAGAAATTAACAAGAT 3940
Db 154 AspPheAsnThrProLeuSerThrLeuAspArgSerThrArgGlnLysValAsnLysAsp 173
QY 3941 ATCCAGGAGTTGAAGTGTGAGCTCTGGACCAACGCGGACCTAATAGATATCTACAGAACTCCC 4000
Db 174 ThrGlnGluLeuAsnSerAlaLeuHisGlnAlaAspLeuIleAspIleTyArgThrLeu 193
QY 4001 CACCCCAATACACAGAAATATACACTCTTCTCAGCATCACATTCACCTTTTAAATTT 4060
Db 4001 CACCCCAATACACAGAAATATACACTCTTCTCAGCATCACATTCACCTTTTAAATTT 4060
```


QY	6213	CAAGGGACACCTAGGA	CTTCTCAAGGAGAACTCAAA	CCACTGATCAAGGAAATAAGAG	6217
DB	783	hrArgSpValLysAspLeuPheLysGluAsnTyrLysProLeuLeuLysGluIleLysGlu	803		
QY	6273	AGGACACAAACAATGGAA	AAACATTCATGCTCACAGATAGTAAGA	ATCAT	6325
DB	803	luAspThrLysLysTrpLysAsnThrProCysSerTrpValGlyArgIleAsnIleValL	823		
QY	6326	AAAAATGCCATCTCCCA	AAAGTAATATAGATTCACTGCTACCCCATCAAGCTACCAT	6385	
DB	823	ysMetAlaIleLeuProLysValIleTyrArgPheAsnAlaIleProIleLysLeuProM	843		
QY	6386	TGACTTTCTCACAGAAT	TGGAAAAACAACCTTTAAATTTTCATATGGAACCAAAAAAAGA	6445	
DB	843	etThrPhePheThrGluLeuGluLysThrThrLeuLysPheIleTrpAsnGln-LysArg	862		
QY	6446	CCCCACAGACCAACAAT	CTTAAGCAAAAAGAACAAAGCTGGAGGTATCATGCTACCT	6505	
DB	863	AlaArgIleAlaLysSerIleLeuSerGlnLysAsnLysAlaGlyGlyIleThrLeuPro	888		
QY	6506	GACTTAAACCTATACT	ATAAGGCTACAGTAAACCAAACTGCATCGTACTGGTACCAAAAC	6565	
DB	883	AspPheLysLeuTyrTyrLysAlaThrValThrLysThrAlaIleTyrTrpTyrGlnAsn	902		
QY	6566	AGATATATAGACCAAT	GGAAACAGACAGACCTCAGAAAATTACACT-GCAATCTACATC	6624	
DB	903	ArgAspIleAspGlnTrpAsnArgThrGluProSerGluIleThrProLeuThrTyrAsn	922		
QY	6625	CATCTGATCTTTGACAA	ACCTCGACAAACAGCAAGCAATGGAAAAAGGATTCCTTATTTAAT	6684	
DB	923	TyrLeuIlePheAspLysProGluLysAsnLysGlnTrpGlyLysAspSerLeuPheAsn	942		
QY	6685	AAATGTTGTTGAAAAA	CTGGCTAGCCATATGCAGAAAGCTGAAACCTGGATCCCTTCCTT	6744	
DB	943	LysTrpCysTrpGluAsnTrpLeuAlaIleCysArgLysLeuLysValaspproPheLeu	962		
QY	6745	ACACCTTATACAAAAG	TAACTCAAGATGAATTAAAGACTTAAATATAAGACATATAAAC	6804	
DB	963	ThrProTyrThrLysIleAsnSerArgTrpIleLysAspLeuAsnValArgProLysThr	982		
QY	6805	ATAAAAACCCA-GAAGA	AAACCTTAGGCAATACCATTCAGATATGGACATGGCCAAAGAC	6863	
DB	983	IleLysThrLeuGluGluAsnLeuGlyIleThrIleGlnAspIleGlyValGlyLysAsp	1002		
QY	6864	TTCATGACTAAACACCA	CAAGCAATGGCAACAAAGCCAAAGCCAAATAGACAAAGTGGGATCTG	6923	
DB	1003	PheMetSerLysThrProLysAlaMetAlaThrLysAlaLysIleAspLysTrpAspLeu	1022		
QY	6924	ATTAACCTATAGAGCT	TCTGCACAGCAAAAAAAACCTGTCAATGATGAGAGTGAACCAAGCAAC	6983	
DB	1023	IleLysLeuLysSerPheCysThrAla-LysGluThrThrIleArgValAsnArgGlnPr	1042		
QY	6984	TACAGATGGGAGNAAT	TTTTTGCATCTATCGATCTGACAAAGGCTAATATATCCAGAGAT	7043	
DB	1042	oThrThrTrpGluThrIlePheThrThrTyrSerSerAspLysGlyLeuIleSerArgIle	1062		
QY	7044	CTACAGAAACTTAAACA	AAATTTTACAAGAAAAA- - -AACAAACCCCTCAAAATATATGGGC	7099	
DB	1062	eTyrAsnGluLeuLysGlnIleTyrLysLysLysThrAsnAsnProIleLysLysTrpAl	1082		
QY	7100	AAAGGATATAGCAGAC	ACTTCTCAAAGAGACATTTATGACGCCAACAAACATATGAA	7159	
DB	1082	AlysAspMetAsnArgHisPheSerLysGluAspIleTyrAlaAlaLysLysHisMetLy	1102		
QY	7160	AAAAACCTCATCATCT	TGTTGAGAGAAATGCAAAACAAACCAACACACAGTGCATACCA	7219	
DB	1102	sLysCysSerSerSerLeuAlaIleArgGluMetGlnIleLysThrThrMetArgTyrHi	1122		
QY	7220	TCATCATCTAGTTAG	TAAGTGTGATCACTAAAAGCTCAGGAAACCAACAAATGCTGGAGAGG	7279	
DB	1122	sLeuThrProValArgMetAlaIleLysLysSerGlyAsnAsnArgCysTrpArgG	1142		
QY	7280	ATGTGGAGAAATAGGA	ACACTTTTCCACTCTGCTGGGAATGTAAATTAAGTTACTTCAACCAT	7339	

polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG0010-ABG3077 represent novel human diagnostic amino acid sequences of the invention.
 Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 1284 AA;

Alignment Scores:
 Pred. No.: 0 Length: 1284
 Score: 5164.00 Matches: 1049
 Percent Similarity: 78.62% Conservatives: 87
 Best Local Similarity: 72.60% Mismatches: 143
 Query Match: 9.87% Indels: 173
 DB: 22 Gaps: 7

US-10-083-853B-2 (1-29921) x ABG14889 (1-1284)

3401 CGGCAAAATAACAGCTAGTGTCAATATGCGAGGATCAATTCACACATAATAATTA 3460
 1 ArgAlaLysSerProAlaAsnIleMetThrGlySerAsnSerHisIleThrLeu 20
 3461 ACCTTAATGTAATGGCTAAATCCCAATTAAGACACACAGCTGGCAATTTGGATA 3520
 21 ThrLeuAsnValAsnGlnLeuAsnSerProIleLysArgHisArgLeuAlaSerTrpIle 40
 3521 AAGAGTCAGACCCATCAGTGTGCTGATTATGAGGAGCCCATCTCACATGAAAGACACA 3580
 41 LysSerGlnAspProSerValCysCysIleGlnGluThrHisLeuMetCysArgpThr 60
 3581 CATAGGCTCAAAATAAGGATGGAGGAAGATTACCAAGTAATGGAACAAACAAAAA 3640
 61 HisArgLeuLysIleLysGlyTrpArgLysIleTrpGlnAlaAsnGlyLysGln---Lys 79
 3641 AAAGCAGGGTGTCAATCCTAGTCTCTGATAAAGACAGCTTTAAACCAACAAAGATCAA 3700
 80 LysAlaGlyValAlaIleLeuValSerAspLysThrAspPheLysProThrLysIleLys 99
 3701 AGAGACAAGAGCCCATTAATGTAAGGCATCAATGGAACAAGAGAGCTAACT 3760
 100 ArgAspLysGluGlyHisTyrMetMetValLysGlySerIleGlnGlnGluLeuThr 119
 3761 ATCTTAATATACATGCCCAATACAGGAGCACCAGATTCAATGAAGCAAGTCTTAGA 3820
 120 IleLeuAsnMetTyrAlaProAsnThrGlyAlaProArgPheIleLysGlnValLeuSer 139
 3821 GACCTACAAAGAGACTTGAATCCACACAATAATAGTGGGAGCTAAATAATAATAGA 3880
 140 AspLeuGlnArgAspLeuAspSerHisThrLeuIleMetGly----- 153
 3881 CACTTTAACCCCTGCAATATTAGGAGATCAATGAGACAGAGAAAATTAACAAGGAT 3940
 154 AspPheAsnThrProLeuSerThrLeuAspArgSerThrArgGlnLysValAsnLysAsp 173
 3941 ATCCAGAGTTGAAGCTCTGGACCAAGCGGACCTAATAGATATCTACAGAACTCCC 4000
 174 ThrGlnGlnLeuAsnSerAlaLeuHisGlnAlaAspLeuIleAspIleTyrArgThrLeu 193
 4001 CACCCTCAAAATCAACAGAAATATACACTCTCTCAGCATCATTTACACCTATTTAAAT 4060
 194 HisProLysSerThrGluTyrThrPhePheSerAlaProHisSerLysIle 213

4061 GACCATGTAATTTAAGTAAACACCTCTCAGCAATGCAAAAGACAGAAATCCTTAACA 4120
 214 AspHisIleLeuGlySerGluAlaLeuLeuSerLysCysLysArgThrGluIleIleThr 233
 4121 AACAGCTCTCAGACTACAGTGCATCTATTAGAACTCAGAATTAAAGAACTCAGCTCAA 4180
 234 AsnTyrLeuSerAspHisSerAlaIleLysLeuGluLeuArgIleLysAsnLeuThrGln 253
 4181 AATCACACACTACATGGAACCTGCAACCTCTCTGAATCAGTACTGGTAAATAAC 4240
 254 SerArgSerThrThrTrpLysLeuAsnLeuLeuAsnAspTyrTrpValHisAsn 273
 4241 AAAATGAAGCAAAATAAAGATGTTCTTTGAACCAATGAGAACAAAGACACAATGTAC 4300
 274 GluMetLysAlaGluIleLysMetPhePheGluThrAsnGluAsnLysAspThrThrTyr 293
 4301 CAGAACTCTGCGGGCATATTAAAGCAGTGTAGAGGGAAATTTATAGCACTAGATGCC 4360
 294 GlnAsnLeu***AspAlaPheLysAlaValCysArgGlyLysPheIleAlaLeuAsnAla 313
 4361 TACAAGAGAAGCAGGAAATATCTAAATAGACACCTTACATCACAAATTAAGAACTA 4420
 314 HisLysArgLysGlnGluArgSerLysIleAspThrLeuThrSerGlnLeuLysGluLeu 333
 4421 GAGAAGAAGACGCAAACTTCAAAAGCTAGCAGAGACAGCAAGAAATACTAAGATCAGA 4480
 334 GluLysGlnGluGlnThrHisSerLysAlaSerArgGlnGluIleThrLysIleArg 353
 4481 GCAGAACTGAAGGATAGACACAAAAGCCCTTCAATAATAATCAATGAATCCAGGAG 4540
 354 AlaGluLeuLysGluIleGluThrGlnLysThrLysIleAsnGluSerArgSe 373
 4541 CTGGTCTTTGAAGCTCAGCAAAAT---AGACCACCTAGACAGACTAATAAGAGAA 4596
 373 rTppPheGluArgIleAsnLysIleAspArgProLeuAlaArgLeuIleLysLys 393
 4597 AAGAGAGAAGAAATCAAGAGATGCAATAAAAAATGATAAGGGGATATCACACCCAGTCC 4656
 393 sArgGluLysAsnGlnIleAspThrIleLysAsnAspLysGlyAspIleThrAspPr 413
 4657 CACAGAAATACAACTATTATCAGAAATATTATAACACCTCTATGCAAAATAACTAGA 4716
 413 OThrGluIleGlnThrIleArgGluTyrTyrLysHisLeuTyrAlaAsnLysLeuG 433
 4717 AATCTAGAGAAATGATAAATCTCGACACATATGTAGCCTGTATGGACCTTGGGG 4776
 433 uAsnLeuGluGluMetAspThrPheLeuAsp----- 443
 4777 ACAGAAACAAAGGGGTGAATGCAGAAATAAAGACAAACAAAGAGTATGTTGGAA 4836
 443 ----- 443
 4837 GTAGGGGTGAGGGGCAACTTGCCTCTAATGGACAAGGGCCCTGAGCTTTACACCCT 4896
 443 ----- 443
 4897 CTCTATTATTAGGCAAAAGAGATACGAGAGGGTGTGGTGGGAAGAGGTACAGCTGTT 4956
 443 ----- 443
 4957 AGGTCCAGAGTAGGCGCTGCAAGACTGCATTCTCAAAACAATAGGCTCTAGATGCCAGT 5016
 443 ----- 443
 5017 AGATAACCTCAAGGAGCCAGTGCCAGGAGTGTATGCCCTCAGCAAAACCTTCTAGGGCAG 5076
 443 ----- 443
 5077 GCACAGAAGTAAGTTTGGCCACCATTTCTGTATTCACGATAAAGAGTTTGTCTTGTATCAA 5136
 443 ----- 443

QY 5137 GTAGCCTCCAGTGGAAATGCTGAGTTGGTCATGATCCCTTTGGCCCTTTTGGCTCCCAAAA 5196
Db 443 ----- 443

QY 5197 CACATACACCTCTCAAGACTAAACACGAGAAAGTCAAAATCCCTGAATATACCAAGTAAC 5256
Db 444 -ThrTyrThrLeuProArgLeuAsnGlnGluValGluSerLeuAsnArgProIleTh 463

QY 5257 AAGTTCTAAATTAAGCAGCAATTAATGATAGCTACCAACCAAAAAAGTCCAGGACCAAGA 5316
Db 463 rGlyAlaGluIleValAlaIleAsnSerLeuSerThrLysLysSerProGlyProAs 483

QY 5317 CGGATTACAGCCAAATTTCTACAGAGGTACAAAGAGAGTGGTACTATTCTCTCGAA 5376
Db 483 pGlyPheThrAlaGluPheTyrGlnArgTyrLysGluGluValProPhePheLeuL 503

QY 5377 ACTATTCCAAAATAAGAA---AATGGGAATCCCTCCCTAACTCATTTTACGAGGCCAGCA 5433
Db 503 sLeuPheGlnSerIleGluLysGlu-GlyIleLeuProAsnSerPheTyrGluAlaSerI 523

QY 5434 TCATCCTGATACCAAAACCTAGCAGTGACACAACAAAAAGAGGAAATTTACAGGCCATAT 5493
Db 523 leIleLeuIleProLysProGlyArgAspThrThrLysLysGluAsnPheArgProIleS 543

QY 5494 CCCTGATCAACATGATGTAATCTCAATATAAATCTGGCAAAACCAATCCAGCAGC 5553
Db 543 erLeuMetAsnIleAspAlaLysIleLeuAsnLysIleLeuAlaAsnGlnIleGlnGln 563

QY 5554 ACATCAAAAGCTTATCTACCATGATCAATGTTGGCTGATCCCTGGGATGCAAGCTGGT 5613
Db 563 isIleLysLysLeuIleHisaspGlnValGlyPheIleProGlyMetGlnGlyTrpp 583

QY 5614 TCAAAATATGCAATCAATAATGTAGGCCATCATATAACAGAACCAATGACAAAAACC 5673
Db 583 heAsnIleArgLysSerIleAsnValIleGlnHisIleAsnArgAlaLysAspLysAsnH 603

QY 5674 ACATGATTATCTCAATAGATGCAGAAAGGCTTTGTCAAAATTCACAGCCCTTCATGC 5733
Db 603 isMetIleValSerIleAspAlaGluLysAlaPheAspLysIleGlnGlnProPheMetL 623

QY 5734 TAAAAATTCCTCAGTAACTAGCTATCGATGGAATGTATCTCAAAATATAAGAGCTATTT 5793
Db 623 euLysThrIleAsnLysIleGlyIleAspGlyMetTyrPheLysIleIleArgAlaIleT 643

QY 5794 ATAC-AAACCCACAGCCAAATATCATATCTAGTGAATGGCAAAACTGGAAGCATTCCTTTGA 5852
Db 643 yrAspLysThrThrAlaAsnIleIleLeuAsnGlyGlnLysLeuGluAlaPheProLeuL 663

QY 5853 GAACCTGCACAGACAGGATGCCCTCTCTCACCACCTCCTATTCAAGATCTATTGGNAG 5912
Db 663 ysThrGlyThrArgGlnGlyCysProLeuSerProLeuLeuPheAsnIleValLeuGluP 683

QY 5913 TTCTGGCCAGGCAATCAGGCAATAGAAAGAAATAAAGGGTATTCAATAGAAAAGAGAGG 5972
Db 683 heLeuAlaArgAlaIleArgGlnGluLysGluIleLysGlyIleGlnLeuGlyLysGluG 703

QY 5973 AAGTCATATCTCTCTGTTTCGACATGACATGTTTGTATATYTTAGAAAACCCCTCGTCT 6032
Db 703 luValLysLeuSerLeuPheAlaAspMetIleValAspLeuGluAsnProIleValS 723

QY 6033 CAGGCCAAAACCTCCTTAAGCTGATAGCAACTTCAGCAAGTCTCAGGACACAAAATCA 6092
Db 723 erAlaGlnAsnLeuLeuLysLeuIleSerAsnPheSerLysLeuSerGlyTyrIleA 743

QY 6093 ATGTGCAAAATATCAAGCATCTTATACGCGCAATAATAGACAACAGAGAGCCAAATCA 6152
Db 743 snValGlnLysSerGlnAlaPheLeuTyrThrAsnAsnArgGlnThrGluSerGlnIleM 763

QY 6153 TGAGTAACCTCATCTACAAATTCCTACAAAGAGAATAAATACCTAGGAATACAACTTA 6212
Db 763 etSerGluLeuProPheThrIleAlaSerLysArgIleLysTyrLeuGlyIleGlnLeuT 783

QY 6213 CAAGGGACACGTAGGAACCTCTTCAAGGAGAACTACAACCCNCTGATCAAGGAAATAAGAG 6272

Db 783 hrArgAspValLysAspLeuPheLysGluAsnTyrLysProLeuLeuLysGluIleLysG 803

QY 6273 AGGACACAAAACAAATGGAAGAACATTCATCTCACAGATAGTAAGATCAT-----G 6325

Db 803 luAspThrLysLysTrpLysAsnThrProCysSerTrpValGlyArgIleAsnIleValL 823

QY 6326 AAAATGCCATACTGCCCAAGTAAATATATAGATTACGTCTACCCCATCAAGCTACCAT 6385

Db 823 ysMetAlaIleLeuProLysValIleTyrArgPheAsnAlaIleProIleLysLeuProM 843

QY 6386 TGACTTTCTTCACAGAATTTGAAAAACAACATTTAAATTTTCATATGTAGAACCAAAAAAGA 6445

Db 843 etThrPhePheThrGluLeuGluLysThrThrLeuLysPheIleTrpAsnGln-LysArg 862

QY 6446 GCCCACAGAGCCAAAGCAATCTTAAGCAAAAAGAACAAAGCTGAGGTATCATCTGCTACT 6505

Db 863 AlaArgIleAlaLysSerIleLeuSerGlnLysAsnLysAlaGlyGlyIleThrLeuPro 882

QY 6506 GACTTAAACCTATATTAAGGCTACAGTAAACCAAACTGCATGGTACTGGTACCAAAAC 6565

Db 883 AspPheLysLeuTyrTyrLysAlaThrValThrLysThrAlaTrpTyrTrpTyrGlnAsn 902

QY 6566 AGATATATAGCAATGGAACAGACAGACCTCAGAAATTTACT--GCAATCTACATC 6624

Db 903 ArgAspIleAspGlnTrpAsnArgThrGluProSerGluIleThrProLeuThrTyrAsn 922

QY 6625 CATCTGATCTTTGACAAACCTGACAAAACAAGCAATGGAAGGATTCCTCTATTTAAT 6684

Db 923 TyrLeuIlePheAspLysProGluLysAsnLysGlnTrpGlyLysAspSerLeuPheAsn 942

QY 6685 AAATGGTGTGGAAAAACTGGCTAGCCATATGCAGAAAGCTGAAAGCTGGATCCCTTCTCT 6744

Db 943 LysTrpCysTrpGluAsnTrpLeuAlaIleCysArgLysLeuLysValAspProPheLeu 962

QY 6745 ACACCTTATACAAAAGTTAACTCAAGATGAATTAAGACITAAATATAAGACATAAACC 6804

Db 963 ThrProTyrThrLysIleAsnSerArgTrpIleLysAspLeuAsnValArgProLysThr 982

QY 6805 ATAAAAACCCCA--GAAGAAAACCTAGGCAATACCATTCAGGATATGGCATCGGCAAGAC 6863

Db 983 IleLysThrLeuGluGluAsnLeuGlyIleThrIleGlnAspIleGlyValGlyLysAsp 1002

QY 6864 TTCATGACTAAACACCAAGCAATGGCAACAAAGCCAAATAGACAACTGGGATCTG 6923

Db 1003 PheMetSerLysThrProLysAlaMetAlaThrLysAlaLysIleAspLysTrpAspLeu 1022

QY 6924 ATTAACCTATAGACTTCTGCACAGCAAAAANAACCTGTCATCAGCTGACACAGCAACC 6983

Db 1023 IleLysLeuLysSerPheCysThrAla-LysGluThrThrIleArgValAsnArgGlnPr 1042

QY 6984 TACAGAATGGAGAAAAATTTTGCATCTATCGATCTGACAAAAGGCTAATATCCAGAGAT 7043

Db 1042 oThrTrpGluThrIlePheThrThrTyrSerSerAspLysGlyLeuIleSerArgIl 1062

QY 7044 CTACGAGAAGCTTAAACAAATTTACAAGAAAA---AACACCCCTCAAAATATGGGC 7099

Db 1062 eTyrAsnGluLeuLysGlnIleTyrLysLysThrAsnAsnProIleLysLysTrpAl 1082

QY 7100 AAAGGATATGAGCAGACACTCTCAAAAGAGACATTTATGCAGCCCAACAAACATATCAA 7159

Db 1082 aLysAspMetAsnArgHisPheSerLysGluAspIleTyrAlaAlaLysLysHisMetLy 1102

QY 7160 AAAAACCTCATCATTTGGTTCGTAGAGAAATCAAAACAAAAACACACAGTGACATCCA 7219

Db 1102 sLysCysSerSerSerLeuAlaIleArgGluMetGlnIleLysThrThrMetArgTyrHl 1122

QY 7220 TCTCATCTAGTTAGATGGTATCACTAAAAAGTCAGGAAACAACAAATGCTCGAGAGG 7279

Db 1122 sLeuThrProValArgMetAlaIleIleLysLysSerGlyAsnAsnArgCysTrpArgGl 1142

QY 7280 ATGTGGAGAAATAGGAACACTTTTCCACTGTTGGTGGAAATGTAATATTAGTTCAACCAT 7339

Db 1142 yCysGlyGluIleGlyThrLeuLeuHisCysTrpTrpAspCysLysLeuValGlnProLe 1162
QY 7340 GTGGAAGACAGTGTGGAGATTCCTTAAGATCTAGAACACAGAAATATCATTTGACCCAGC 7399
Db 1162 uTrpLysSerValTrpArgPheLeuArgAspLeuGluLeuGluIleProPheAspProAl 1182
QY 7400 ATCCCATCTAGTATATACCCAAAGAAATATAATCATCTCTATTATAAGACACATG 7459
Db 1182 aileProLeuLeuGlyIleTrpProAsnAspTrpLysSerCysCysTrpLysAspThrCy 1202
QY 7460 CACATATGTTTATTGAGCACTCATCATATAGCAAGACTTGGAAACCAACCAATG 7519
Db 1202 sThrArgMetPheIleAlaLeuPheThrIleAlaLysThrTrpAsnGlnProLysCy 1222
QY 7520 TCATCAGTGTAGTGTGATAAGAAACATGCGCATATATACCATGAAATACATATGC 7579
Db 1222 sProThrMetIleAspTrpIleLysLysMetTrpHisIleTrpMetGluTrpTyrAl 1242
QY 7580 ACCATAAAG-GATGAGTTCATGCTCTTTCAGAGATATGATGAGCTGGAAACCAT 7638
Db 1242 aAlaIleLysAsnAspGluPheMetSerPheValGlyThrTrpMetLysLeuGluThrI 1262
QY 7639 CATCTCAGCAACTAACAAGAAACAGAAACCAACACACATGTTCTCACTGTGAAG 7698
Db 1262 eileLeuSerLysLeuSerGlnGluGluLysThrLysHisArgIlePheSerLeuIle 1282
QY 7699 TGGGAGT 7705
Db 1282 yGlyAsn 1284

RESULT 7

ABGI4545
ID ABGI4545 standard; Protein; 1726 AA.

XX ABGI4545;

XX DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #14536.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR N-PSDB; AAS78732.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity

XX PS Claim 20; SEQ ID No 44904; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG3037 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 1726 AA;

Alignment Scores:

Pred. No.: 0 Length: 1726
Score: 5086.00 Matches: 1154
Percent Similarity: 61.73% Conservative: 141
Best Local Similarity: 55.00% Mismatches: 298
Query Match: 9.72% Indels: 514
DB: 22 Gaps: 31

US-10-083-853B-2 (1-29921) x ABGI4545 (1-1726)

QY 1210 CATTCCTCCAGTACCGTACCGTTCATCTC----- 1239
Db 8 HisProProAlaGluTrpProArgHisValLeuLeuLeuMetAlaAspSerLysAsnGlu 27
QY 1240 -----ACTGGGACTGGTTAGACATTGGGTGGCGAGCCACGGAAGGTGACTGAA--- 1287
Db 28 LeuGlnAsnMetAsnThrGlyGlnThrCysGlyThrArgGlnThrLeuLysSerLysLeu 47
QY 1288 GCAGGTGGGTGTCCCTCAGCCCGAAGTGCAGAGGGGTGGGGATCTCTCTCCGCCA 1347
Db 48 GlnGlyTrpAsnLysPro-----GlyIleGlnSerAspPro 59
QY 1348 GCCAAGGAAGCCATGACAGACTGTACAGGAGGAATGCTGACACTCTAGTCAGATACGTG 1407
Db 60 GlySerGlnGlyHisProArg-CysLeuLys----- 69
QY 1408 CACTTTTCCCATAGTCTTTGCAACTGGCAGACAGGAGATTTCCCGCAGTGCCTATGCCA 1467
Db 70 -ThrPheProLeuValAlaGlnSerSer-----ThrGlnGluArgAsnHis 84
QY 1468 C-----CAGGGCCCTGGGTTTCAAGCACAAACTGGCGGCCATV----- 1507
Db 84 sValMetGluLysAlaValLeuLysSerGlnAsnTrpAlaIleIleGluGluPheIlePh 104
QY 1508 -----TGGACAGACACCGAGCTAGCCGCCACAGTTATTTTTCATATCCCGCTGG 1557
Db 104 rLysIleGlyTrpPheProAspTyrSerGlyLysValLysTyrLeuSerProLysSe 124
QY 1558 CGCTTGAATGCCAAGACAGACACCATTCACCTCCAGGGATCCAAAGTGGTCTGCTCAG 1617
Db 124 rSer-----CysThrThrValLeuGlnValThrHisPheSe 136
QY 1618 T---GGTCCCACCCCATGGAGCCAGCTA---GCTAAGATCCAC----- 1657
Db 136 rLeuPhePheHisProAsnProAspGluLeuValSerLysIleLysValTrpSerLysHis 156
QY 1658 -----TGCTTGAAATTCCTCGCCAGCACAGCAGTC 1689
Db 156 sArgLeuTyrGlnAsnAsnSerGluAlaPheLeuGluValGlnIleProGluProLysCy 176
QY 1690 TGAGATTCA-----CCTGGG---ATGCTTGACCTGGG 1718
Db 176 sGluValTyrSerMetArgThrMetGlyArgGlnProGlyThrAlaMetAspLeuAs 196

Db 675 uGlnArgAspLeuAspSerAsnThrIleIleThrGly-----AspH 689
QY 3886 TAACACCCCACTGCCAATATTAGGCAGATCAATAGACAGACAAAATTAACAGGATATCCA 3945
Db 689 eAsnThrProLeuSerThrLeuAspArgSerMetArgGlnLysValAsnLysAspIleGI 709
QY 3946 GGAGTTGAAGTCTGGACCAAGCGGACCTAATAGATATCTACAGAACTCCGCCACCC 4005
Db 709 nGluLeuAsnSerAlaLeuHisGlnAlaAspLeuIleAspIleYrArgThrLeuHisPr 729
QY 4006 CAATCAACAGAAATATACACTCTCTCAGCATCATCATACCTATTTTAAATGACCA 4065
Db 729 oLysSerThrGluAsnThrPhePheSerAlaProHisHisThrYrSerLysIleAspHI 749
QY 4066 TGTAATTTTAAGTAAACACTCCTCAGCAATGCAAAAGACAGAAATCCTTAACAAACAG 4125
Db 749 sIleValGlySerLysAlaLeuLeuSerLysCysLysArgLysGluIleIleThrAsnCy 769
QY 4126 TCTCTCAGACTACAGTCAATCTATTATTAGAACTCAGAATTAAAGAACTCACTCAAAATCA 4185
Db 769 sLeuSerAspHisSerAlaIleLysLeuGluLeuArgIleLysLysLeuProGlnAsnCy 789
QY 4186 CACAACACTACATGGAACCTGAAACCTCCTCCTGACTACTGGGTAAATAACAAAT 4245
Db 789 sSerThrIleThrLysLeuAsnAsnLeuLeuAsnAspYrTrpValHisAsnGluMe 809
QY 4246 GAAGCAAAATAAAGATGTTCTTTGAAACCAATGAGAAACAGACACAATGTACACGAA 4305
Db 809 tLysAlaGluIleLysMetLeuPheGluThrAsnGluAsnLysAspThrThrYrGlnAs 829
QY 4306 TCTCTGGGCATATTATAAGCAGTGTGTAGAGGAAATTTATAGCACTAGATGCTACAA 4365
Db 829 nLeuTrpAspThrLeuLysAlaValCysArgGlyLysPheIleAlaLeuAsnAlaHisGI 849
QY 4366 GAGAAAGCAGCAATATCTAAATAGACACCTTACATCACAATTAAGAACTAGAGAA 4425
Db 849 yArgLysGlnGluArgSerLysIleAspThrLeuThrSerGlnLysGluLeuAlaLy 869
QY 4426 GAAAGACAAACAATCAAAAGCTAGCAGAGACAGAAATAACTAAGATCAGACGAGA 4485
Db 869 sGlnGluGlnThrHisSerLysAlaSerArgArgGlnGluIleThrLysIleArgAlaGI 889
QY 4486 ACTGAGAGATAGACACAAAAGCCCTTCAATAAATCAATGAATCCAGGAGCTGGT 4545
Db 889 uLeuLysGluIleGluThrGlnLysThrLeuGlnLys-IleAsnGluSerArgSerTrp 909
QY 4546 TTTTGAAGATCAGCAAAAT---AGACCACCTAGACAGACTAATAAGAGAAAGAG 4601
Db 909 hePheGluArgIleAsnLysIleAspArgProLeuAlaArgLeuIleLysLysArg 929
QY 4602 AGAAGAAATCAAGAGATGCAATAAAAAATGATAAGGGGATATCACCACCGATCCACAG 4661
Db 929 LuLysAsnGlnIleAspThrIleLysAsnAspLysGlyAspIleThrThrAspProThrG 949
QY 4662 AAATACAACTATTATCAGAAATATTATAACACCTCTATGCAAAATAACTAGAAATC 4721
Db 949 luIleGlnThrThrIleArgGluYrYrLysHisLeuYrAlaAsnLysLeuGluAsnL 969
QY 4722 TAGAAGAAATGGATAAATTTCTTGACACATATGTAGCCTGTATGGACCTTGGGGACAGA 4781
Db 969 euGluGluMeCaspThrPheLeuAsp----- 977
QY 4782 ACAAAGGGGTGAATGAGAAATAAAGACAAAAGAGTATGTTTGAAGTAGG 4841
Db 977 ----- 977
QY 4842 GGTACGGGGCACTTGCTCTAATGGACAAAGGGCCCTGAGCTTTACACCCTCTGTA 4901
Db 977 ----- 977
QY 4902 TTTATTAGCAAAAGAGATACCGAGGGTGTGGAGAGAGTGTAGTGTAGTGC 4961
Db 977 ----- 977

QY 4962 CAGAGTAGGCTGCAAGACTGCATTCTCAACAATAGGCTCTAGATGTCCAGTAGATA 5021
Db 977 ----- 977
QY 5022 ACCTCAAGGAGCCAGTGCCAGGAGTGATGGCCCTCAGCAAACTTCTTAGGCAGGCACA 5081
Db 977 ----- 977
QY 5082 GAAGTAAGTTTGCCCAACATTCTGTATTCCAGATAAAGCTTGTGTTGTATCAAGTAGC 5141
Db 977 ----- 977
QY 5142 CTCAGTGAATGCTGAGTTGGTCATGATCCCTTTGGCCTTTTGGCTCCCAAAACACAT 5201
Db 978 -----ThrT 979
QY 5202 ACACCTCTCAAGACTAAACAGGAGAAAGTCAAACTCCTGAATATACCATACAGTAACAGTT 5261
Db 979 yrThrLeuProArgLeuAsnGlnGluValGluSerLeuAsnArgProIleThrGlys 999
QY 5262 CTAAATTTGAAGCAGTAATTGATAGCTTACCACCAACAAAAGTCCAGGACGACGAT 5321
Db 999 erGluIleValAlaIleIleAsnSerLeuProThrLysLysSerProGlyProaspGlyP 1019
QY 5322 TCAGAGCCAAATTTTACCAGAGGTACAAAGAGAGCTGGTACTATTCTCTGAAACTAT 5381
Db 1019 heThrAlaGluPheYrGlnArgYrLysGluGluLeuValProPheLeuLeuLysLeuP 1039
QY 5382 TCCAAAATAATGAA---AATGGGAATCTCCTCACTCACTTTTACGAGGCGCAGCATATC 5438
Db 1039 heGlnSerIleGluLysGlu-GlyIleLeuProAsnSerPheYrGluThrSerIleIle 1058
QY 5439 CTGATACCAAAACCTAGCAGTGACACACAAAAGAGAAATTTTCAGGCCCATATCCCTG 5498
Db 1059 LeuIleProLysProGlyArgAspThrThrLysLysGluAsnPheArgProIleSerLeu 1078
QY 5499 ATGAACATTGTGTGAAATCCTCAATAAATACTGCGAAACCAACCAATCCAGCAGCACATC 5558
Db 1079 MetAsnIleAspAlaLysIleLeuAsnLysIleLeuAlaAsnArgIleGlnHisIle 1098
QY 5559 AAAAGCTTATCTACCATGATCAAGTTGGCTCATCTCCCTGGATGTCAGGCTGGTTCAA 5618
Db 1099 LysLysLeuIleHisHisAspGlnValGlyPheIleProGlyMetGlnGlyTrpPheAsn 1118
QY 5619 ATATGCAATCAATAATGTAGGCATCACATAACACCAACCAATGACAAAACACACATG 5678
Db 1119 IleArgLysSerIleAsnValIleGlnHisIleAsnArgAlaLysAspLysAsnHisMet 1138
QY 5679 ATTATCTCAATAGATGCAAGAAAGCCCTTTGTCAAAATTCACAGCCCTTCATGCTAAAA 5738
Db 1139 IleIleSerIleAspAlaGluLysAlaPheAspLysIleGlnGlnProPheMetLeuLys 1158
QY 5739 ATCTCAGTAACTAGGTATCGATGGAATGTATCTCAAAATATCAAGACTATTATATAC- 5797
Db 1159 ThrLeuAsnLysLeuGlyIleAspGlyThrYrPheYrIleIleArgAlaIleYrAsp 1178
QY 5798 AAACCCAGCAGCAATATCATACTGAATGGGCAAAACCTGGAAGCATTCCTTTGAGAAGT 5857
Db 1179 LysProThrAlaAsnIleIleLeuAsnGlyGlnLysLeuGluAlaPheProLeuLysThr 1198
QY 5858 GGCACAGACAGGATGCCCTCTCTCACCCTCTCTATTCAAGATATATTGAGAGTCTG 5917
Db 1199 GlyThrArgGlnGlyCysProLeuSerProLeuLeuPheAsnIleValLeuGluValLeu 1218
QY 5918 GCCAGGGCAATCAGCAATAGAAATAAAGGTATTCAATAGAAAGAGAGAGATC 5977
Db 1219 AlaArgAlaIleArgGlnGluLysGluIleLysGlyIleGlnLeuGlyLysGluVal 1238
QY 5978 ATATTGCTCTGTTGGCAGATCATCTTTTATATTTAGAAAACCCCTCTCTCAGGC 6037
Db 1239 LysLeuSerLeuPheAlaAspMetIleValYrLeuGluAsnProIleValSerAla 1258

Alignment Scores:

Pred. No.: 0 Length: 2030
Score: 5067.50 Matches: 1100
Percent Similarity: 65.15% Conservative: 128
Best Local Similarity: 58.36% Mismatches: 233
Query Match: 9.68% Indels: 412
DB: 22 Gaps: 14

US-10-083-853B-2 (1-29921) x ABG24249 (1-2030)

QY	2188	TCCACGACCTGCGACGAGCGGCTGACTCTTAGAGGAAAGTAACAACAAAGG	2247	3146	ATATCCACATCTTTAAAGAAAATAATTTTCAACCCAGAATTTTCATATCCAGCCAAACCA	3205
Db	446	SerAsnArgProAlaAlaGluGlyProValValLeuGluGlyLysLeuThrThr	463	725	-----	725
QY	2248	AATAGTATCAACATTAACAAAAGGACATCCACTCAGAGACCCCATCTGAAGGTCAACAA	2307	3206	AGCTTCTTAAGTGAAGGAGAAAATAAAATCTCTACAGAGAAGCAAAATGCTGACAGATTTT	3265
Db	464	-----ArgLysAspIleThrThrGluAsnProSerValHisHis	477	725	-----	725
QY	2308	CATCAAGACCAAGGTAAATAAACCAGATGGGAAAC-CAGTGCAGAAACACT	2366	3266	TGTCACCACCGGCTGCTTACAGAGCTCTCTGAGGAAGCACCACCATGGAAGGAAC	3325
Db	478	HisGlnArgProLysValAspLysThrThrLysMetGlyLysGlnSerArgLysThr	497	725	-----	725
QY	2367	GAAATTCAAAACCAAGACTCTCTCTCAACCAAGGATCAACACTCTCGCCAGCA	2426	3326	AACTGGTACCACCGACTGCAAAAACATCCAAAATTTGTAAGACCATTGCTATGAAGA	3385
Db	498	GlyAsnSerLysLysGlnSerAlaSerProProProLysGluCysSerSerProAla	517	725	-----	725
QY	2427	AGGGAACAAACAGATGAGATGAGTTGAGGAATTCACAGAAAGTAGGCTTCAGAAGG	2486	3386	AAGTCATCAACTAAGGGGCAAAATAACACGCTAGTGTCTATATGCGCAGGATCAAAATTC	3445
Db	518	ThrGluGlnSerTrpThrGluAsnAspPheAspGluLeuArgGluGluGlyPheArgArg	537	725	-----	725
QY	2487	TGGGTAAATCAAACTCTCCGAGCTAAAGGAGCATGTTCTAACCCAAATGCAAGGAAGCT	2546	3446	CACATAATAATATTAACTTTAAATGTAAATGGCTAAATTTCCCAATTTAAAGACACAGA	3505
Db	538	-----SerAsnTyrSerGluLeuGlnGluGluGlnThrLysGlyLysGluVal	554	725	-----	725
QY	2547	AAGAACCTTGAAGAAGGTAGATGAATTCCTAACTAGATATATCAGTGTACAGAGAAC	2606	3506	CTGGCAAAATTGGATAAAGAGTCAAGACCACATCAGTGTGTATTTCAGGAGGCCCATCTC	3565
Db	555	GluAsnPheGluLysAsnLeuAspGluThrArgIleThrArgIleThrAsnThrLysCys	574	725	-----	725
QY	2607	ATAATGACTGATGAGCTGAAAACGCAAGACAAAGAACTTCATGAACATACACAAGC	2666	3566	ACATGAAAAGACACACATAGGCTCAAAATAAAGGGATGGAGGAAGATTTTACCNAAGTAAT	3625
Db	575	LeuLysGluLeuMetGluLeuLysAlaLysAlaArgGluLeuArgGluGluCysArgSer	594	725	-----	725
QY	2667	TTCATAGCAATCGATCAACGAGAAAGGATATCAGTGTGAAGATCAAAATTAAT	2726	3626	GGAAACAAAAAAGACAGGGTTCGAATCTCTGATCTCTGATAAAACAGACTTTAA	3685
Db	595	LeuArgSerArgCysAspGlnLeuGlnGluArgGluSerValMetGluAspGluMetAsn	614	726	-----AspPheLys	728
QY	2727	AAAAGAAAGTGAGACACAAGATTACAGAAAAAGAGTCAAAAAGAAACAAACAGCCCTC	2786	3686	CCAACAAAGATCAAAAGAGACAAAGAGCCATTACATAATGTAAGAGCATCAATGGAA	3745
Db	615	GluMetLysArgAlaValLysPheThrLysLysArgIleLysArgAsnGluGlnThrLeu	634	729	ProThrThrIleLysArgAspLysGluGlyHisTyrIleMetValLysGlySerIleGln	748
QY	2787	CAAGAAATATGGACTATGTGAAGACCAAAATCTACATTTGATTGGTCTCCCAAGT	2846	3746	CAAGAAGACTTAATCTCTTAATATACATGACCCCAATACAGGACCCAGATTCATA	3805
Db	635	GlnGluIleTrpAspTyrValLysArgProAsnLeuArgLeuIleGlyValProLysSer	654	749	GlnGluGluLeuThrIleLeuAsnIleTyrAlaProAsnThrGlyAlaProArgPheIle	768
QY	2847	GATGGGCAATGGAATCAAGTTGGAACACTCTTCAGGGTATTATCCAGAGAAATTC	2906	3806	AAGCAAGTTCTTAGAGACCTTACAAAGAGACTTTGACTCCACACAATAATAGTGGAGTC	3865
Db	655	AspGlyGluAsnGlyThrLysLeuGluAsnThrLeuGlnAspIleIleGlnGluAsnPhe	674	769	LysGlnValLeuSerAspLeuGlnArgAspLeuAspSerHisThrLeuIleMetGly---	787
QY	2907	CCCAT-CTATCAGGCGAGCCCAACATTCAAATTCAGGAATATGGAGACACCATAAAGA	2965	3866	TAAATAATAATAGACACTTTAACCCCACTGCCCATAATATTAGGAGATCAATGAGACAG	3925
Db	675	ProAsnLeuAlaArgGlnAlaAsnIleGlnIleGlnIleGlnArgThrProGlnArg	694	788	-----AspPheAsnThrProLeuSerThrLeuAspArgSerThrArgGln	802
QY	2966	TACTCTCAGAGAACCAATCCCAAGACACATAATCTTCAGATTCACCAAGGTGAAATG	3025	3926	AAAATAACAAGGATATCCAGGAGTTGAAGTCTGACCCAGCCAGCCGCTCAATAGAT	3985
Db	695	TyrSerSerArgArgAlaThrProArgHisIleIleValArgPheThrLysValLysMet	714	803	LysValAsnLysAspThrGlnGluLeuAsnSerAlaLeuHisGlnAlaAspLeuIleAsp	822
QY	3026	AAGGAAAAATGTTAAGGCGACCCAGAGAAAGGTGGGTTTACCCCAAGAGGAGCCA	3085	3986	ATCTACAGAACTCCCAACCCCAAAATCAACAGAAATATACACTCTCTTCAGCATTCACATTC	4045
Db	715	LysGluLysMetLeuArgAlaAlaArgGluLys-----	725	823	IleTyrArgThrLeuHisProLysSerThrGluTyrThrPhePheSerAlaProHisHis	842
QY	3086	ATCAGACTAACAGCGGATCTCCCGCAGAAACCTTACAGCCAGAGAGTGAAGGCCA	3145	4046	ACCTATTTTAAATTTGACCATGTATTTTAACTTAAACACTCTCTCAGCAAAATGCAAAAGA	4105
Db	725	-----	725	843	ThrTyrSerLysLysIleAspHisIleLeuGlySerLysAlaLeuLeuSerLysCysLysArg	862
				4106	ACAGAAATCTTAACAAACAGTCTCTCAGACTACAGTGCATCTATTATTAGAACTCAGAAAT	4165
				863	ThrGluIleIleThrAsnTyrLeuSerAspHisSerAlaIleLysProGluLeuArgIle	882
				4166	AAGAACTCACATCAAAATCACAACTACATGAACTGAACTGAACTGCTCTGCTGATGAC	4225
				883	LysLysLeuThrGlnAsnCysSerThrThrTrpLysLeuAsnLeuLeuLeuAsnAsp	902
				725	-----	725

QY 4226 TACTGGGTAAATAACAAATGAAGCAAAATAAGATGTTCTTTTGAACCAATGAGAAC 4285
Db TyrTrpValHisAsnGluMetLysAlaGluIleLysMetPheGluThrAsnGluAsn 922
QY 4286 AAAGACACAATGTACCAGATCTCTGGGCGATATTTAAAGCAGTGTGTAGAGGAATTT 4345
Db LysAspThrThrTyrglnAsnLeuTrpAspThrPheLysAlaValCysArgGlyLysPhe 942
QY 4346 ATACACTAGATGCTCAACAGAGAAAGCAGCAATATCTAAATAGACACCTTACATCA 4405
Db IleAlaLeuAsnAlaHisLysArgLysGlnGluArgSerLysIleAspThrLeuThrSer 962
QY 4406 CAATTAAAGAACTAGAGAAAGAGCAAAATTCAAAGCTAGCAGAGACAGAA 4465
Db GlnLeuLysGluLeuGluLysGlnGluThrHisSerLysAlaSerArgArgGlnGlu 982
QY 4466 ATACTAGATCAGACGAGAACTGAAGGATAGAGACAGCAAAAGCCCTTCAATTAAT 4525
Db IleThrLysIleArgAlaGluLeuLysGluIleGluThrGlnLysThrLeuGlnLys-I 1002
QY 4526 CAATGAATCCAGGAGCTGTTTTTGAAGAGATCAGCAAAAT---AGACCACCTAGACAG 4581
Db eAsnAspSerArgSerTrpPhePheGluLysIleAsnLysIleAspArgProLeuAlaAr 1022
QY 4582 ACTAATAAAGAAAGAGAGAGAAATCAAGAGATGCAATAAAAAATGATAAGGGGA 4641
Db gProLysLysLysLysArgGluLysAsnGlnIleAspAlaIleLysIleAspLysGlyAs 1042
QY 4642 TATCACCACCGATCCCAGAGAAATACAACTATTATCAGAGAAATATTAACACCTCTA 4701
Db pileThrAsnProThrGluIleGlnThrIleArgGluTyrlLysArgLeuTy 1062
QY 4702 TGCAAATAAAGTACAGAAATCTAGAGAAATGGATAAAATTCCTGGACACATATGTAGCCTG 4761
Db rAlaAsnLysLeuAlaAsnLeuGluGluMetAspLysPheAsp 1077
QY 4762 TATGGACCTTGGGGGACAGAACAAAGGGGTGAATGCAGAAATAAGACAAAGACAAA 4821
Db 1077 ----- 1077
QY 4822 AGAGTATGTTTGAAGTAGGGGTGAGGGGCAACTTGCCCTCTAATGGACAGGCCCTGA 4881
Db 1077 ----- 1077
QY 4882 GCTTTACACCCTCTGTATTATTATAGCAAAAGATAGCAGAGGGGTGAGTTGGAAG 4941
Db 1077 ----- 1077
QY 4942 AAGAGTCAGCTGTAGGTCAGAGTAGGCCGTGCAAGACTGCATTCCTCAAAACAATAGGC 5001
Db 1077 ----- 1077
QY 5002 TCTAGATGTCCAGTAGATACCTCAAGGAGCCAGTGCCAGGGAGTGATGGCCCTCAGCA 5061
Db 1077 ----- 1077
QY 5062 AACCTTCTAGGCAGGCACAGAGTAAGTTTGCCACATTCGTGATTACAGATAAAGCT 5121
Db 1077 ----- 1077
QY 5122 TTGCTGTTGNTCAAGTAGCCTCCAGTGGAAATGCTGAGTTGGTCATGATCCCTTTGGCCT 5181
Db 1077 ----- 1077
QY 5182 TTTTGGCTCCCAAAACACATACCCCTCAAGACTAAACAGAGAAAGTCAAAATCCCT 5241
Db -----ThrTyrlThrLeuProArgLeuAsnGlnGluValGluSerLe 1092
QY 5242 GAATATACCAAGTAACAGTTCTAAATTTGAAGCAGTAATGATAGCCTACCAACCAAAA 5301
Db uAsnArgProLysThrGlyAlaGluIleValAlaIleIleAsnSerValProThrLysLy 1112
QY 5302 AAGTCCAGGACCAAGGATTCACGCCAAATTTCTACAGAGGTACAAAGAGAGCTGGT 5361

Db sSerProGlyProAspGlyPheThrAlaLysPheTyrglnSerTyrlLysGluGluLeuVa 1132
QY 5362 ACTATTCTCTCTGAACATATTCCCAAAAATAGAA---AATGGGAATCCTCCCTCACTCAT 5418
Db lProPheLeuLeuLysLeuPheGlnSerIleGluLysGlu-ArgIleLeuProAsnSerP 1152
QY 5419 TTTACGAGGCAGCATCATCTGATACCAAAACCTAGCAGTGACACAAACAAAGAGAA 5478
Db heTyrgluAspSerIleIleLeuIleProLysProGlyArgAspThrThrLysLysGluA 1172
QY 5479 ATTTACGCCCATATCCTCGATGAACATTTGATGTGAAATCTCTCAATAAATACTGGCAA 5538
Db snPheArgProIleSerLeuMetAsnIleAspAlaLysIleLeuAsnLysIleLeuAla 1192
QY 5539 ACCAAATCCAGCACACATCAAAAGCTTATCTACCATGATCAAGTTGGCGCTCATCCCTG 5598
Db snGlnIleGlnGlnHisLysLysLeuIleHisAspGlnValGlyPheIleProG 1212
QY 5599 GGATGCAAGGCTGCTTCAAAATATGCAAAATCAATAAATGTAGGCATCACATAACAGAA 5658
Db lYMetGlnGlyTrpSerThrIleHisLysSerIleAsnValIleGlnHisIleAsnArgT 1232
QY 5659 CCAATGACAAACACACATGATTATCTCAATAGATGAGAAAGCGCTTGTCTCAAAATTC 5718
Db hrLysAspLysLysHisMetIleIleSerIleAspAlaGluLysAlaPheAspLysIleG 1252
QY 5719 AACAGCCTTCATCTAAATCTCAGTAACTAGGTATGATGATGATGATGATGATGATGAT 5778
Db lGlnProPheMetLeuLysThrLeuAsnLysLeuGlyIleAspGlyThrTyrlLysLysI 1272
QY 5779 TAATAAGAGCTATTATATAC-AAACCCACAGCAATATCATACTGAATGGGCAAAACTGG 5837
Db leIleSerAlaIleTyrlAspLysProThrAlaAsnIleIleLeuAsnGlyGlnLysLeuG 1292
QY 5838 AAGCATTCCTTTGAGAACTGGCAACAGACAGATGCCCTCTCTCACCACCTCTATATCA 5897
Db luAlaLeuProLeuLysThrGlyThrArgGlnGlyCysProLeuSerProLeuLeuPheA 1312
QY 5898 AGATCTATTGGAAGTTCTGGCCAGGCAATCAGGCAATAGAAAGAAATAAGGTATTC 5957
Db snIleValLeuGluValLeuAlaArgAlaIleArgGlnGluLysGluIleLysGlyIleG 1332
QY 5958 AAATAGAAAGAGAGAAAGTATGTTCTGTTTGCAGATGACATGTTGTATATATTAG 6017
Db lLeuGlyLysGluGluValLysLeuSerLeuPheAlaAspMetIleValTyrlLys 1352
QY 6018 AAAACCCCTGCTCTAGGCCCAAAACTCCTTAAGCTGATAGCAACTTCAGCAAGTCT 6077
Db luAsnProIleValSerValGlnAsnLeuHisLysLeuIleSerAsnPheSerLysValS 1372
QY 6078 CAGACACAAAAATCAATGTGCAAAATCAACAGCATTTTATAGCCCAATATAGACAAA 6137
Db erGlyTyrlLysIleAsnValGlnLysSerGlnAlaPheLeuTyrlThrAsnAsnArgGluT 1392
QY 6138 CAGAGACCAATCATCAGTCACTCATTCACAATTTGCTACAAAGAGATAAAATACC 6197
Db hrGluAsnGlnIleMetSerGluLeuProPheThrIleAlaSerLysArgIleLysTyrlL 1412
QY 6198 TAGGAATACAACTTACAGGACAGCTAGGAACCTTCAAGGAACTACAAACCACTGA 6257
Db euGlyIleHisLeuThrArgAspValLysAspLeuPheLysGluAsnTyrlLysProLeuL 1432
QY 6258 TCAAGGAATAAGAGAGACACAAACAAATGGAATAATCCATGCTCACAGATGATAA 6317
Db euAsnGluIleLysGluAspThrAsnLysTyrlLysAsnIleProCysSerTrpValGlyA 1452
QY 6318 GAATCAT-----GAAATGCCATAGTCCCAAGTAATATATAGATTCAGTCTCACC 6370
Db rgIleAsnIleValLysMetAlaIleLeuProLysValIleTyrlArgPheAsnAlaIleP 1472
QY 6371 CCATCAGCTACCATTCACCTTCTTTCACAGATTTGGAAACCAACTTAAATTTTCATAT 6430
Db

Db	1472	roileuysLeuProMetThrPhePheThrGluLeuGluLysThrThrLeuLysPheIleT	1492
QY	6431	GGAAACCAAAAAAGAGCCACAGACCAAGCAAACTTTAAGCAAAAAAGCAAAAGCTGGA	6490
Db	1492	rpAsnGln-LysArgAlaHisIleAlaLysSerIleLeuSerGlnLysAsnLysAlaGly	1511
QY	6491	GGTATCATGCTACCTGACTTAAACTATACTATAGGTACAGTACCAAAAACCTGCATGG	6550
Db	1512	GlyIleMetLeuProAspPheLysLeuThrHisLysAlaThrValThrLysThrAlaTrp	1531
QY	6551	TACTGGTACCAAAACAGATATATACCAATCGACAGACAGACACCTCAAGAAATTACA	6610
Db	1532	TyrTrpTyrGlnAsnArgAspIleAspGlnTrpAsnLysThrGluProSerGluIleThr	1551
QY	6611	CT-GCAATCATCATCTGATCTTGCACAACTGCACAAAACCAAGCAATGGAAAAG	6669
Db	1552	ProHisIleTyrAsnTyrLeuIlePheAspLysProGluAsnLysGlnTrpGlyLys	1571
QY	6670	GATTCCTTATTAATAATGGTGTGGAAAACTGGCTAGCCATATGCAGAAAGCTGAAA	6729
Db	1572	AspSerLeuPheAsnLysTrpCysTrpGluAsnTrpLeuAlaIleCysArgLysLeuLys	1591
QY	6730	CTGGATCCCTCTCTACACCTTATACAAAGTTAACTCAAGATGAATTAACACCTTAAT	6789
Db	1592	LeuAspProPheLeuThrProTyrThrLysIleAsnSerArgTrpIleLysAspLeuAsn	1611
QY	6790	ATAAGACATAAAACCAACCAACCA-GAAGAAAACTAGGCAATACCATTCAGGATATG	6848
Db	1612	ValArgProLysThrIleLysThrLeuGluLysLeuGlyIleThrIleGlnAspIle	1631
QY	6849	GACATGGGCAAGACTTCATGACTAAACACCAAGCAATGGCAACAAAGCCCAATAA	6908
Db	1632	GlyIleGlyLysAspPheMetSerLysThrProLysAlaMetAlaThrLysAspLysIle	1651
QY	6909	GACAAAGTGGATCTGATTAACATATAGAGCTTCTCAGACAGCAAAAACCTCATCAG	6968
Db	1652	AspLysTrpAspLeuLysLysSerPheCysThrAla-LysGluThrIleLeu	1671
QY	6969	AGTGAACAGCAACCTCAGAAATGGAGAAAATTTTGCATCTATCGATCTGACAAAGG	7028
Db	1671	gValAsnArgGlnProThrLysTrpGluGluPheAlaThrTyrSerSerAspLysGln	1691
QY	7029	CTAATATCCAGAGTCTAGGAAGAATTAAACAATTTACAAGAAAAA-----ACAACCC	7084
Db	1691	yLeuIleSerArgIleTyrAsnGlyLeuLysGlnIleTyrLysLysThrAsnAsnPr	1711
QY	7085	CGTCAAAATATGGCAAGGATATGAGCAGACACTTCTCAAAAGAGACATTATGTCAGC	7144
Db	1711	oileLysTrpPalalalysAspMetAsnArgHisPheSerLysGluAspIleTyrAlaAl	1731
QY	7145	CAACAAACATATGAAAAAACCTCATCATCTGCTAGAGAAATGCAAAAACAAAC	7204
Db	1731	alysLysHisValLysLysCysSerSerSerLeuAlaIleArgGluMetGlnIleLysTh	1751
QY	7205	CACAGTGACATACCATCTCATGCTAGTATAGATGGTGATCACTAAAGAGTCAGGAACAA	7264
Db	1751	rThrMetArgTyrHisLeuThrProValArgMetAlaIleLysLysSerGlySerAs	1771
QY	7265	CAAATCGCTGGAGAGGTGGAGAAATAGGACACCTTTCCACTGTGGTGGGATGTAA	7324
Db	1771	nArgCysTrpArgGlyCysGlyGluIleGlyThrLeuLeuHisCysTrpTrpAspCysLys	1791
QY	7325	ATTAGTTCAACCATTTGTGAAGACAGTGTGGAGATTCCTTAAGGATCTAGAACCAAGAT	7384
Db	1791	sleuValGlnProLeuTrpLysSerValTrpArgPheLeuArgAspLeuGluLeuGluI	1811
QY	7385	ATCATTTGACCGACCAATCCCATTTACTGAGTATATACCAAGAGGAATATAATCTCTA	7444
Db	1811	ePropPheAspProAlaIleProLeuLeuGlyIleTyrProAsnAspTyrLysSerCysCys	1831
QY	7445	TTATAAGACACATGCACACATATGTTT-----AT 7474	
Db	1831	styLysAspThrCysThrLeuSerLeuProGluIleArgAsnCysProThrLysSerVa	1851
QY	7475	TGCAGCACTGATCAACAATAGCAAGACTTGGAAACCAACCAATGCTCCATCATGATAGA	7534
Db	1851	IProValProPheSerValAspGluThrGlnTyrHisValGluAspThrSerProValse	1871
QY	7535	CTGGATAAAGAAAAATGG-----CACATATACACCATGAATACTATGACAGCCATA--	7586
Db	1871	rTrpGlnValLysValLeuGlyLysHisIleHisGlnGlySerSerTyrGluAspValTr	1891
QY	7587	-----AAAAGGATGAGTTTCATGCTCTTGCAGAGATATGATGAAGCTGGAACCATC	7639
Db	1891	pLeuProGlnArgMetGlnGly-LysLeuThrAsnArgLysAsp-----IleH 1907	
QY	7640	ATTCTCAGCAAACTAACACACAGAACAGAAACCAACACCATGTT-----CTCAC	7693
Db	1907	IstThrLysAsnProSerValHisHisHisGlnArgProLysValAspLysThrThrL	1927
QY	7694	GTAAGTGGGAGTTGAACATGAGACACATGCACACAGGAGGGAACATCACACACAG	7753
Db	1927	ysMetGlyLysLysGlnAsnArgLysThrGlyAsnSerLysLysGlnSerAlaSerProp	1947
QY	7754	GTCCT 7758	
Db	1947	roPro 1948	
RESULT 9			
ID	ABG10129	standard; Protein; 2764 AA	
XX	ABG10129;		
DT	13-FEB-2002	(first entry)	
XX	Novel human diagnostic protein #10120.		
XX	Human; chromosome mapping; gene mapping; gene therapy; forensic;		
KW	food supplement; medical imaging; diagnostic; genetic disorder.		
XX	Homo sapiens.		
XX	WO200175067-A2.		
PD	11-OCT-2001.		
PF	30-MAR-2001; 2001WO-US08631.		
PR	31-MAR-2000; 2000US-0540217.		
PR	23-AUG-2000; 2000US-0649167.		
XX	(HYSE-) HYSEQ INC.		
PA	Drmanac RT, Liu C, Tang YT;		
PI	WPI; 2001-639362/73.		
DR	N-PSDB; AAS74316.		
XX	New isolated polynucleotide and encoded polypeptides, useful in		
PT	diagnostics, forensics, gene mapping, identification of mutations		
PT	responsible for genetic disorders or other traits and to assess		
PT	biodiversity -		
XX	Claim 20; SEQ ID No 40488; 103pp; English.		
XX	The invention relates to isolated polynucleotide (I) and		
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,		
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome		
CC	and gene mapping, and in recombinant production of (II). The		
CC	polynucleotides are also used in diagnostics as expressed sequence tags		
CC	for identifying expressed genes. (I) is useful in gene therapy techniques		
CC	to restore normal activity of (II) or to treat disease states involving		
CC	(II). (II) is useful for generating antibodies against it, detecting or		
CC	quantitating a polypeptide in tissue, as molecular weight markers and as		
CC	a food supplement. (II) and its binding partners are useful in medical		


```
Db 415 rAlaIleLysLeuGluLeuArgIleLysAsnLeuAsnGlnSerArgSerThrThrTrpLys 435
QY 4201 ACTGAACAACCTGCTCTGAATGACTACTGGTAAATAACAAATGAAGCAAAATAAA 4260
Db 435 sLeuAsnAsnLeuLeuLeuAsnAspTyrTrpValHisAsnGluMetLysAlaGluIleLys 435
QY 4261 GATGTTCTTTGAACAAGTGAACAACAAAGACAAATGTACAGAACTCTGGGCGATATT 4320
Db 455 sMetPhePheGluThrAsnGluAsnLysAspThrThrTyrGlnAsnLeuTrpAspAlaPh 475
QY 4321 TAAAGCAGTGTGTAGAGGGAATTTATAGCAGCTAGATGCTTACAGAGAAAGCAGGAAT 4380
Db 475 eLysAlaValCysArgGlyLysPheIleAlaLeuAsnAlaHisLysArgLysGlnGluAr 495
QY 4381 ATCTAAATAGACACCTTAAACATCACAATTAAGAAGAACTAGAGAAAGAAAGACAAACA 4440
Db 495 gSerLysIleAspThrLeuThrSerGlnLeuLysGluLeuGlyGlnGluGlnThrHI 515
QY 4441 TTCAAAAGCTAGCAGAGACAAGAAATAACTAAGATCAGAGCAGAACTGAAGGAGATAGA 4500
Db 515 sSerLysAlaSerArgArgGlnGluIleThrLysIleArgAlaGluLeuLysGluIleGI 535
QY 4501 GACAAAAAGCCCTTCAATAATCAATCAATCCAGGAGCTGGTTTTTGAAGAATCA 4360
Db 535 uThrGlnLysThrLeuGlnLys-IleAsnGluSerArgSerTrpPheGluArgIleA 555
QY 4561 GCAAAAT----AGACCACATAGACAGACTAATAAGAAGAAAGAGAGAGAAATCAAGAG 4616
Db 555 snLysIleAspArgProLeuAlaArgLeuIleLysLysLysArgGluGluAsnGlnIleA 575
QY 4617 ATGCAATAAAAAATGATAAGGGGATATCACCCAGATCCACAGAAATACAAACTATTA 4676
Db 575 spThrIleLysAsnAspLysGlyAspIleThrThrAspProThrGluIleGlnThrThrI 595
QY 4677 TCAGAGATATTAACACCTCTATGCAAAATAACTAGAGAAATCTAGAGAAATGGATA 4736
Db 595 leArgGluTyrLysHisLeuTyrAlaAsnLysLeuGluAsnLeuGluGluMetAspL 615
QY 4737 AATTCCTGGACACATATGTAGCCTGTGTAGCACCTTGGGGGACAGACAAAGGGGTGAA 4796
Db 615 yspHeLeuasp----- 618
QY 4797 TGCAGAAATAAAGACAAAGACAAAAGAGTAGTATTTTGAAGTAGGGGTACAGGGGCAACT 4856
Db 618 ----- 618
QY 4857 TGCCCTCTAATGGACAGGGCCCTGAGCTTTACACCACCCCTCTGTATTATTAGGCAAAAG 4916
Db 618 ----- 618
QY 4917 AGATAGCGAGAGGGTGAGTTGGAAGAAGAGGTGAGTGTGTAGGTCCAGAGTAGGCCGTGCA 4976
Db 618 ----- 618
QY 4977 AGACTGCATTCCTCAAAACATAGGCTCTAGATGTCCCGTAGATACCTCAAGGAGCCAG 5036
Db 618 ----- 618
QY 5037 TGCCAGGGAGTGTAGTGGCCCTCAGCAAAACCTTCTAGGGCAGGCACAGAGTAAGTTTGCCC 5096
Db 618 ----- 618
QY 5097 ACATTCCTGTATTACGATAAACAGTTTGCTGTGTGTGTTGATCAAGTAGCCTCCAGTGAATGCT 5156
Db 618 ----- 618
QY 5157 GAGTTGGTATGATFCCCTTTGGCCCTTTTGGCTFCCCAAAACATACACCCCTCTCAAGAC 5216
Db 619 -----ThrTyrThrLeuProArgL 625
QY 5217 TAAACAGAGAGAGTCAATCCCTGAATATACCAAGTAAACAAGTTCTAAATTTGAGACAG 5276
Db 625 euAsnGlnGluValGluSerLeuAsnArgProIleThrGlyAlaGluIleValAlaI 645
```

```
QY 5277 TAATTGATAGCTTACCAACCAAAAAAGTCCAGGACGAGGATTTCACAGCCAAATTCT 5336
Db 645 leIleAsnSerLeuProThrLysLysSerProGlyProAspGlyPheThrAlaGluPheT 665
QY 5337 ACCAGAGGTACAAGAGAGAGCTGTACTATTCTCTTGAACATATTCCCAAAAATAGAA- 5395
Db 665 yrGlnArgTyrLysGluGluValProPheLeuLeuLysLeuPheGlnSerIleGluL 685
QY 5396 --AATGGGAATCCCTCACTACTATTTTACGAGCGCAGCATCATCTGATACCAAAACCT 5453
Db 685 ysglu-GlyIleLeuProAsnSerPheTyrGluAlaSerIleIleLeuIleProLysPro 704
QY 5454 AGCAGTGACACAACAAAAAGAGAAATTTCCAGGCCATATCCCTGTGATGAACATTGTGTG 5513
Db 705 GlyArgAspThrThrLysLysGluAsnPheArgProIleSerLeuMetAsnIleAspAla 724
QY 5514 AAAATCCTCAATAAATACTGGCAACCAATCCAGCAGCACATCAAAAAGCTTATCTAC 5573
Db 725 LysIleLeuAsnLysIleLeuAlaAsnArgIleGlnGlnHisIleLysLysLeuIleHis 744
QY 5574 CATGATCAAGTTGCGCTCATCCCTGGGATGCAAGGCTGTTCAAAATATCAATCAATA 5633
Db 745 HisAspGlnValGlyPheIleProGlyMetGlnGlyTrpPheAsnIleArgLysSerIle 764
QY 5634 AATGTAGGCCATCACATAAACAAGACCAATGACAAAAACCAATGATTATCTCAATAGAT 5693
Db 765 AsnValIleGlnHisIleAsnArgAlaLysAspLysAsnHisMetIleIleSerIleAsp 784
QY 5694 GCAGAAAAGCCCTTTGCAAAATTCACAGCCCTTCATGTGTAATAAATCTCAGTAACATA 5753
Db 785 AlaGluLysAlaPheAspLysIleGlnGlnProPheMetLeuLysThrLeuAsnLysLeu 804
QY 5754 GSTATCGATGAATGTATCTCAAAATAAAGAGCTATTATTATAC-RAACCCACAGCCAAT 5812
Db 805 GlyIleAspGlyThrTyrPheLysIleIleArgAlaIleTyrAspLysProThrAlaAsn 824
QY 5813 ATCATACTGAATGGGAAAAAAGCTTCCCTTTGAGAACTGTCACAAAGACACAGGA 5872
Db 825 IleIleLeuAsnGlyLysLysLeuGluAlaPheProLeuLysThrGlyThrArgGlnGly 844
QY 5873 TGCCCTCTCTCACCATCTTATTCAGATACTATTGGAAGTTCTGGCAGGCGCAATCAGG 5932
Db 845 CysProLeuSerLeuLeuPheAsnIleValLeuGluValLeuAlaArgAlaIleArg 864
QY 5933 CAATAGAAAGAAAATAAGGCTATTCAAAATAGAAGAGAGAGAACTCATATTGTCTCTTT 5992
Db 865 GlnGluLysGluIleLysGlyIleGlnLeuGlyLysGluGluValLysLeuSerLeuPhe 884
QY 5993 GCAGATGACATGTTGTATATTTAGAAAAACCCCATCTCTCAGGCCAAAAAAGCTCTTAAG 6052
Db 885 AlaAspAspMetIleValTyrLeuGluAsnProIleValSerAlaGlnAsnLeuLys 904
QY 6053 CTGATTAAGCACTTCAGCAAAAGTCTCAGGACACAAAATCAATGTGCAAAAATCACAGCA 6112
Db 905 LeuIleSerAsnPheSerLysValSerGlyTyrLysIleAsnValGlnLysSerGlnAla 924
QY 6113 TTCTTTATGCCCAATATACAGAAACAGAGACAAATCATGAGTGAACCTCATTCACACA 6172
Db 925 PheLeuTyrThrAsnAsnArgGlnThrGluSerGlnIleMetSerGluLeuProPheThr 944
QY 6173 ATTGCTCAAAAAGAAATAAATACCTAGGAATACAACTTACAAAGGACAGCTGAACTC 6232
Db 945 IleAlaSerLysArgIleLysTyrLeuGlyIleGlnLeuThrArgAspValLysAspLeu 964
QY 6233 TTCAAGGAGAACTACAAACACCTGATCAAGAAATTAAGAGAGACACAAACAATGAAA 6292
Db 965 PheLysGluAsnTyrLysProLeuLeuLysGluIleLysGluAspThrAsnLysTrpLys 984
QY 6293 AACATTCCTGCTCACAGATAGTAAGATCAT-----GAAATGCCATCTGCCCAA 6345
Db 985 AsnIleProCysSerTrpValGlyArgIleAsnIleValLysMetAlaIleLeuProLys 1004
```


XX
PT
PT
PT
PT
XX

New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity -

Claim 20; SEQ ID No 39511; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG0010-ABG3037 represent novel human diagnostic amino acid sequences of the invention.
Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1824 AA;

Alignment Scores:

Pred. No.:	0	Length:	1824
Score:	4989.50	Matches:	1101
Percent Similarity:	67.01%	Conservative:	134
Best Local Similarity:	59.74%	Mismatches:	260
Query Match:	9.54%	Indels:	355
DB:	22	Gaps:	28

US-10-083-853B-2 (1-29921) x ABG09152 (1-1824)

QY	2490	GTAAATACAACTCTCCGAGCTAAAGGAGCATGTTCTTAACCCAA-----	2534
DB	131	ilelleLeuAsnGlyGlnLysLeuGluAlaPheThrLeuLysThrGlyThrArgGlnGly	150
QY	2535	TGCAAGGAGCTAAAGAACCTTCAAAAAGGTAGATGAATTCCTACTAGATAATCACT	2594
DB	151	CysProLeuSerProLeuLeuPheAsnIleValLeuGluValAlaIleArg	170
QY	2595	GTAGAGAAACATAAATGACCTGATGGAGCTCAAAAACGCAAGACAACTTCATGAA	2654
DB	171	GlnGluLysGluIleLysGlyLe-----	178
QY	2655	GCATACACAGCTTCAATAGCCAAATCGATCAAGCAGAAAGGATATCACTGATGAA	2714
DB	179	-----GlnLeuGlyLysGluGluValLysLeuSerLeuPheAla	191
QY	2715	GATCAAAAT-----	2723
DB	192	AspAspMetIleValTyrLeuGluAsnProIleIleSerAlaGlnAsnLeuLysLeu	211
QY	2723	-----	2723
DB	212	IleGlyAsnPheSerLysValSerGlyTyrThrIleAsnValGlnLysSerGlnAlaPhe	231
QY	2724	-----AATAAAGAAAGTCAGAACAG-----	2750
DB	232	LeuTyrThrAsnAsnArgGlnThrGluSerGlnIleMetSerGluLeuProPheThrIle	251
QY	2751	ACAGAAAAGACTGAA-----AGAAACAACAACCCCAAGATTAATGGACATATGTG	2807
DB	252	AlaSerLysArgIleLysTyrLeuGlyIleGlnLeuThrArgAspValLysAspPhePhe	271

QY	2808	AAA-----AGACCAAAATCTACATTTGTTGGTGTCCCCCAAAAGTGATGGGAGAAAT	2858
DB	272	LysGluAsnTyrLysProLeuLeuAsnGluIle-----LysGluAsp	285
QY	2859	GGAATCAAGTTGGAAACACTC-----	2882
DB	286	-LysAsnLysTyrLysAsnIleProCysSerTrpValGlyArgIleAsnIleValLysMe	305
QY	2883	CAGGGTATTCAGGAGAAATTCCTCCATCTATCAGGCGAGGCAACATTCAAATTCAGG	2942
DB	305	talalleLeuProLysArgThr-ProLeuLeuSerLeuProSerThrLeu---TyrGluG	324
QY	2943	AAATATGGAGA-----ACACCAATAAGATACCTCTCGAGAGAA	2981
DB	324	luLeuTrpArgLeuArgSerTyrPheProLeuCysProGlnLeuTyrAsnMetGluGlnA	344
QY	2982	CAATCCCAAGACACATAATCTTCAGA-----TTCACCAAGTTGAAATGA	3026
DB	344	snThrThrAspTrpValIleTyrLysGluGlnCysIleTrpLeuThrAlaLeuGluThrG	364
QY	3027	AGGAAAAATGTTAAGGCGCAGCAGAGAGAAAGTTGGTTACCCACAAGGGAAGCCAA	3086
DB	364	lyGluSerLysSerThrAlaPro-----AlapheGlyLysGlyTyrProM	379
QY	3087	TC-----AGACTAACA-----GCGGATCTCCCGGCAAAA	3116
DB	379	etalaglutyrillelleTyrGlyArgLeuThrProHisMetAlaGlyTyrSerSerGlu	399
QY	3117	CCCTACAGCCAGAGAGAGAGTGGGCCAATATTCACATTCCTTAAGAAAAATAATTTTC	3176
DB	399	hr-LysLeuProGluGluArgSerGly-----SerAsnIleCysCys	412
QY	3177	NAACCCAGATTTTCATATCCAGCCAAACCAAGCTTCTTAAGTAGAGAGAAATAAATCT	3236
DB	413	SerPro--IleSerAlaValLeuGlnProLeuLeuLeuLeuProArgGlnThrGlySerG	432
QY	3237	CTACAGAG---AAGCAAAATGCTGACAGATTTTGTCAACCCAGGCTCCCTTACAGAG	3293
DB	432	lyValAspLeuGlnGlnThrProThrAspLeuGlnLeuArgValLeuThrValArg-	451
QY	3294	CTCCTGAGGAGAGCACCACATCGAAGAAAGCAACTGGTACCAG---CACTGCAAAAAC	3350
DB	452	-----LysThrAsnLysGlnLysGlyHisProHisGlnAsnProIleCysThrS	468
QY	3351	ATCCCAAAATGTAAAGACCATTCATGCTATGAAGAAAGTGCACTCAACTAACGGCAAAAT	3410
DB	468	erPro-----	469
QY	3411	AACCGACTAGTGTCTAATGGCAGGATCAATTCACACATAATAATATTAACTTAATG	3470
DB	470	-----SerSerLysThrLysGlySerAsnSerHisIleThrIleLeuThrLeuAsnV	487
QY	3471	TAAATGGGTAAATTCCTCCCAATTAAGACACAGACTGCGCAATTCGATTAAGAGTCAAG	3530
DB	487	alAsnGlyLeuAsnAlaProIleLysArgHisArgLeuAlaAsnTrpIleLysSerGlnA	507
QY	3531	ACCCATCAGTGTGTATTTCAGAGGCCATCTCACATGAAAGACACACATAGGCTCA	3590
DB	507	sProSerValSerCysIleGlnGlnThrHisLeuThrCysArgAspThrHisArgIleL	527
QY	3591	AAATAAGGGATGGAGGAAGATTTACCAAGTAATGGAACAAAAAAGGAGGGG	3650
DB	527	ysileLysGlyTrpArgGluIleTyrGlnAlaAsnGlyLysGln---LysLysAlaGlyV	546
QY	3651	TTGCAATCTAGTCTCTGATAAACAGACTTTAAACCAACAAGATCAAAAGACAGAAAG	3710
DB	546	alalleLeuValSerAspLysThrAspPheLysProThrLysIleLysArgAspLysG	566
QY	3711	AGGGCCATACATAATGGTAAGGCATCAATGGAACAAGAGCTAACTATCTCTAAATA	3770
DB	566	luGlyHisTyrMetMetValLysGlySerIleGlnGlnGluLeuThrThrLeuAsnI	586


```
QY 3771 TACATGCCCAATACAGGAGCACCAGATTATAAAGCAAGTCTTTAGACACCTACAAA 3830
Db |||:|||||
QY 586 leYrAlaProAsnThrGlyAlaProArgPheIleLysGlnValLeuArgAspLeuGlnA 606
Db |||:|||||
QY 3831 GAGACTTTGACTCCACACATAATAGTGGAGTCTAAATAATAATAGACACTTTACA 3890
Db |||:|||||
QY 606 rgAspLeuAspSerHisThrLeuIleMetGly-----AspPheAsnT 620
QY 3891 CCCCACTGCCAATATAGGCAGATCAATGAGACGAAAAATTAACAGGATATCCAGGAGT 3950
Db |||:|||||
QY 620 hrProLeuSerThrLeuAspArgSerThrArgGlnValAsnLysAspIleGlnAspL 640
QY 3951 TGAACCTGAGCTCGACCAAGCGACCTAAATAGATATCTACAGAACTCCCAACCCCAAT 4010
Db |||:|||||
QY 640 euAsnSerAlaLeuHisGlnValAspLeuIleAspIleYrArgThrLeuHisProLys 660
QY 4011 CAACAGATATACACTCTTCTCAGCATCACATTACACCTATTTAAATTCACCATGTAA 4070
Db |||:|||||
QY 660 erThrGluYrThrPhePheSerAlaLeuHisHisIleYrSerLysIleAspHisIleV 680
QY 4071 TTTTAAGTAAACACTCTCTCAGCAAAATGCAAAAGACAGAAATCTTAACAAACAGTCTCT 4130
Db |||:|||||
QY 680 aiGlySerLysAlaLeuLeuSerLysYrLysThrThrGluIleIleThrAsnCysLeu 700
QY 4131 CAGACTACAGTGCATCTATTATTAGAACTCAGAATTAAGAACTCCTCAAAATCACACAA 4190
Db |||:|||||
QY 700 erAspHisSerAlaIleLysLeuGluLeuArgIleLysLysLeuThrGlnAsnArgSert 720
QY 4191 CTACATGGAACACTGACAACTGCTCCTGATGATGACTACTGGTAAATTAACAAATGAAG 4250
Db |||:|||||
QY 720 hrThrTrpLysLeuAsnAsnLeuLeuAsnAspTyrTrpValHisAsnLysMetLysA 740
QY 4251 CAAAAATAAAGATGTTCTTCAACCAATGAGACAAAGACACAAATGTAACAGAAATCTCT 4310
Db |||:|||||
QY 740 laGluIleAsnThrLeuPheGluThrAsnGluAsnLysAspThrThrYrGlnAsnLeuT 760
QY 4311 GGGGCATATTTAAAGCAGTGTGTAGAGGAAATTTATAGCACTAGATGCCCTACAGAGAA 4370
Db |||:|||||
QY 760 rpAspThrPheLysAlaValCysArgGlyLysPheIleAlaLeuAsnAlaHisLysArgL 780
QY 4371 AGCAGGAATATCTAAATACACACTTAACATCACATTAATTAAGAACTAGAGAAAG 4430
Db |||:|||||
QY 780 ysGlnGluArgSerLysIleAspThrLeuThrSerGlnLeuLysGluLeuGluLysGlnG 800
QY 4431 AGCAACAAATTCAAAAGCTAGCAGAGACAGAAATAAATCAAGTACAGACAGAACTGA 4490
Db |||:|||||
QY 800 luGlnThrHisSerLysAlaSerArgArgGlnGluIleThrLysIleArgAlaLysLeuL 820
QY 4491 AGGAGATAGACACAAAGCCCTTCAAATAATCAATCAATCAATCAAGAGCTGTTTTTT 4550
Db |||:|||||
QY 820 ysGluIleGluThrGlnLysThrLeuGlnLys-IleAsnGluSerArgSerTrpPhePhe 839
QY 4551 GAAAGATCAGCAAAATAGACCA---CTAGACACACTAATAAGACAGAAAGAGAGAG 4606
Db |||:|||||
QY 840 GluGlnIleAsnLysIleAspArgLeuLeuAlaArgLeuIleLysLysLysArgGluLys 859
QY 4607 AATCAAGAGATGCAATAAAAATGATAAGGGGATATCACACCGATCCCAAGAAATA 4666
Db |||:|||||
QY 860 AsnGlnIleAspThrIleLysAsnAspLysGlyAspIleThrThrAspProThrGluIle 879
QY 4667 CAACATATTATCAGAGATATTTAAACACCTCTATGCAATAAATAGAAATCTAGAA 4726
Db |||:|||||
QY 880 GlnThrThrIleArgGluYrYrLysHisLeuYrAlaAsnLysLeuGluAsnLeuGlu 899
QY 4727 GAAATGATTAATTCCTGGACACATATGTAGCCCTGTATGGACCTTGGGGGACAGACAA 4786
Db |||:|||||
QY 900 GluMetAspLysPheLeuAsp----- 906
QY 4787 AGGGGGTGAATGCAGAAATAAAGACAAAGACAAAGAGATGTTTGGAGTAGGGGTCA 4846
Db |||:|||||
QY 906 ----- 906
QY 4847 GGGGGCAACTTGGCTCTTAATGGACAGGGCCCTGAGCTTTTACACACCCCTCTGTATTAT 4906
Db |||:|||||
```


Db	1128	gAlaileArgGlnGluLysGluLeuLysGlyIleGlnLeuLysGluLeuValLysLe	1148
QY	5983	GTCTCTGTTGACAGATGTTTGTATATTTAGAAAACCCCATCGTCTCAGGCCAAAA	6042
Db	1148	userLeuPheAlaAspMetIleValTyLeuGluAsnProIleValSerAlaGlnAs	1168
QY	6043	ACTCCTTAGCTGATAGCAACTTCAGCAAAAGTCTCAGGACACAAATAATCAATGTGCAAAA	6102
Db	1168	nLeuLeuLysLeuIleSerAsnPheSerLysValSerGlyTyLeuLysIleAsnValGlnLy	1188
QY	6103	ATCACAGCATCTTATACGCCAATAATAGACAAACAGAGACCAATCATGAGTGAAT	6162
Db	1188	sSerGlnAlaPheLeuTyThrAsnAsnArgGlnThrGluSerGlnIleMetSerGluLe	1208
QY	6163	CTCATTACAAATTCCTACAAAGAGAATAAAATACCTAGGAATACAACTTACAAGGGACAC	6222
Db	1208	uProPheThrIleAlaSerLysArgIleLysTyLeuGlyIleGlnLeuThrArgSpVa	1228
QY	6223	GTAGGAACCTTCAAGGAGAACTACAAACCACTGATCAAGGAATAAAGAGAGGACACAAA	6282
Db	1228	lLysAspLeuPheLysGluAsnTyLysProLeuLeuLysGluIleLysGluAspThrAs	1248
QY	6283	CAAAATGGAACCAATCCATGCTCACAGATAGTAAGAATCAT-----CAAAATGCCAT	6335
Db	1248	nLysTrpLysAsnIleProCysSerTrpValGlyArgIleAsnIleValLysMetAlaIl	1268
QY	6336	ACTGCCCAAGTAATATAGATTACAGTGTACGCCCATCAAGCTACCATGACTTCTT	6395
Db	1268	eLeuProLysValIleTyArgPheAsnAlaIleProIleLysLeuLeuIleThrPhePh	1288
QY	6396	CACGAATTTGGAACCAACTTAAATTTTCATATGGAACCAACCAAAAGAGCCACAGAG	6455
Db	1288	eThrGluLeuGluTyThrLeuLysPheIleTrpAsnGln-LysArgAlaArgIleA	1308
QY	6456	CCAAGCAATCTTAAGCAAAACCAACAAAGCTGGAGGTATCATCTACCTGACTTAAAC	6515
Db	1308	lAlysSerIleLeuSerGlnLysAsnLysAlaGlyGlyIleThrLeuProAspPheLysL	1328
QY	6516	TATACATTAAGCTACAGTAACCAAACTGCATGGTACTGGTACCAACCAAGATATAG	6575
Db	1328	eutyTyLysAlaThrValThrLysThrAlaIrpTyTrpTyGlnAsnArgAspIleA	1348
QY	6576	ACCAATGGACAGACAGACCTCAGAAATTC-ACTGCAATCTACATCCATCTGATCT	6634
Db	1348	spGlnTrpAspArgTrpGluProSerGluIleMetProHisIleTyAsnTyLeuIleP	1368
QY	6635	TTCACAAACCTGCAAAAACCAACCAATGGAAGAGGATCCCTTATTATAAATGGTGT	6694
Db	1368	heAspLysProGluLysAsnLysGlnTrpGlyLysAspSerLeuPheAsnLysTrpCys	1388
QY	6695	GGAAAACCTGGCTAGCCATATGC---AGAAAGCTGAAACTGGATCCCTTCTTACACCTT	6751
Db	1388	rpGluAsnTrpLeuAlaIleCysArgArgLysLeuLysLeuAspProPheLeuThrProT	1408
QY	6752	ATACAAAGTTAACTCAAGATGAATTAAGACTTAATAATPACATATAAACCAATAAAAA	6811
Db	1408	yrThrLysIleAsnSerArgTrpIleLysAspLeuAsnValArgProLysThrIleLysT	1428
QY	6812	CCCA-GAAGAAACCTAGGCAATACCATTCAGATATGGACATGGCGCAAGAGCTTCATGA	6870
Db	1428	hrLeuGluGluAsnLeuGlyAsnThrIleGlnAspIleGlyMetGlyLysAspPheMets	1448
QY	6871	CTAAAAACCAAAAGCAATGGCAACCAAAAGCCAAAATAGACAGTGGGATCTGATTAAC	6930
Db	1448	erThrThrProLysAlaMetAlaThrLysAlaLysIleAspLysTrpAspLeuIleLysL	1468
QY	6931	TATAGAGCTTCTGCACAGCAAAAAAACTGTATCATCAGATGAACAGCAACCTACAGAA	6990
Db	1468	eulysSerPheCysThrAla-LysGluThrThrIleThrValAsnArgGlnProThrGlu	1487
QY	6991	TGGGAGAAATTTTCCATCTATCGATCTGACAAAGGCTAATATATCCAGAGATCTACGAA	7050
Db	1488	TrpGluLysIlePheThrIleTyProSerAspLysGlyLeuIleSerArgIleTyLys	1507
QY	7051	GAACCTTAAACAAATTTTACAAGAAAAA-----AACAAACCCCTCAAAATATATGGCAAGGAT	7106
Db	1508	GluLeuLysGlnIleTyLysLysArgThrAsnAsnProIleAsnLysTrpValLysAsp	1527
QY	7107	ATGAGCAGACACTTCTCAAAAGAGACATTTATGACGCCCAACAAACATATGAAAAAAC	7166
Db	1528	MetAsnArgHisPheSerLysGluAspIleTyThrAlaLysArgHisMetLysLysCys	1547
QY	7167	TCATCATCATGCTGCTTACAGAAATGCAAAACCAACACAGTGCATACCATCTCATG	7226
Db	1548	SerSerSerLeuAlaIleArgGluMetGlnIleLysThrLeuArgTyHisLeuIle	1567
QY	7227	CTAGTTAGAAATGTGATCACTAAAAAGTCAGGAAACAAACAACTGCTGAGAGGATGTGGA	7286
Db	1568	ProValArgMetAlaIleIleLysLysSerGlyAsnAsnArgCysTrpArgGlyCysGly	1587
QY	7287	GAATAGGAACACTTTTCCACTGTTGGTGGGAATGTAAATAGTTCAACCATGTGTGAAG	7346
Db	1588	GluIleGlyThrLeuLeuHisCysTrpTrpAsnCysLysLeuValGlnProLeuTrpLys	1607
QY	7347	ACAGTGTGGAGATTCCTTAAGGATCTAGAACCAAGAAATATCATTTGACCCAGCAATCCCA	7406
Db	1608	SerValTrpArgPheLeuArgAspLeuGluLeuGluIleProPheAspProAlaIlePro	1627
QY	7407	TTACTGAGTATATACCCAAAGGAATATAAATCTCTTATTATAAGACACATGCACACAT	7466
Db	1628	LeuLeuGlyIleTyProLysAspTyLysSerCysCysTyLysAspThrCysThrArg	1647
QY	7467	ATGTTTATTCACACATGATCAATAGCAAAAGACTTGGAACCAACCAATGTCCATCA	7526
Db	1648	MetPheIleAlaAlaLeuPheThrIleAlaLysAlaTrpAsnGlnAlaLysCysProThr	1667
QY	7527	GTGATAGACTGGATAAGAAACATGSCACATATACACCATGAATACTATGCAGCCATA	7586
Db	1668	MetIleAsnTrpIleLysLysMetTyPargIleTyThrMetGluTyTyAlaAlaIle	1687
QY	7587	AAAAG-GATGAGTTCATGCTCTTTGACAGAGATATGGATGAAGCTGGAACCAACCATCTCTC	7645
Db	1688	LysAsnAspGluPheMetSerPheValGlyThrTrpMetLysLeuGluIleIleLeu	1707
QY	7646	AGCAAACTACACAAAGACAGAAACCAACCAACCAACATGTCTCCTACTGTA	7696
Db	1708	SerLysLeuSerGlnGluLysThrLysHisCysIleLeuSerLeuIle	1724
RESULT 11			
ID	ABG08447	standard; Protein; 2921 AA.	
AC	ABG08447;		
XX			
DT	13-FEB-2002 (first entry)		
XX			
DE	Novel human diagnostic protein #8438.		
XX			
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;		
KW	food supplement; medical imaging; diagnostic; genetic disorder.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200175067-A2.		
XX			
PD	11-OCT-2001.		
XX			
FF	30-MAR-2001; 2001WO-US08631.		
PR	31-MAR-2000; 2000US-0540217.		
PR	23-AUG-2000; 2000US-0649167.		
XX			
PA	(HYSE-) HYSEQ INC.		
XX			
PI	Drmanac RT, Liu C, Tang YT;		
XX			

DR WPI; 2001-639362/73.
DR N-PSDB; AAS72634.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 38806; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2921 AA;

Alignment Scores:
Pred. No.: 0 Length: 2921
Score: 4902.00 Matches: 1065
Percent Similarity: 63.19% Conservative: 111
Best Local Similarity: 57.23% Mismatches: 223
Query Match: 9.37% Indels: 469
DB: 22 Gaps: 13

US-10-083-853B-2 (1-29921) x ABG08447 (1-2921)

QY 2343 GGGAAACACAGTCGACAGAACACTCAAAATCCAAAACACAGAACTCTCTTCAACCA 2402
DB 2 GlyLysLysGlnSerArgLysThrGlyAsnSerLysLysGlnSerAlaSerProPro 21
QY 2403 AAGATCACAACTCTCCAGCAGGAGGAGAAACACAGATGGAGATGAGTTGAGGAA 2462
DB 22 LysLysArgSerSerProAlaThrGluGlnSerTrpThrGluAsnAspPheAspGlu 41
QY 2463 TTGACAGAGTAGGCTTCAGAGGTGGGTAAATACAACTCTCCGAGCTAAAGAGCAT 2522
DB 42 LeuArgGluGluGlyPheArg-----SerAsnTyrSerGluLeuArgGluGlu 58
QY 2523 GTTCTAACCCCAATCAAGAGAGCTAAGAACCTTGAAAAAGGTAGATGATTCCTAACT 2582
DB 59 IleGlnThrLysGlyLysGluValGluAsnTyrGluLysSerLeuGluGluCysIleThr 78
QY 2583 AGAATAATCAGTGTAGACAGAACACATAATACCTGATGGAGCTGAAAAACGCAAGCAA 2642
DB 79 ArgIleThrAsnThrGluAsnCysLeuLysGluLeuMetGluProLysThrLysAlaArg 98
QY 2643 GAATTCATGAAGCATACACAGCTTCAATAGCAAAATCGATCAAGCAGAAAGAGGATA 2702
DB 99 GluLeuArgLysGluCysArgSerLeuArgSerArgCysTyrGlnLeuValGluArgVal 118
QY 2703 TCAGTGATGAAGTCAAAATTAATAAGAAAGTGAAGACAAAGATTACAGAAAAAGA 2762
DB 119 SerAlaMetGluAspGluValAsn----- 126
QY 2763 GTGAAGAAACAAACAAAGCCTCCAGAAATTATGGGACTATGTGAAGAACCAATCTA 2822

DB 127 -----Glu--Met----- 128
QY 2823 CATTTGATTGTTGTCCTCCCAAGTATGGGGAGATGAATCAAGTTGGAACACACTCTT 2882
DB 128 ----- 128
QY 2883 CAGGTATTATCCAGAGAAATTTCCCATCTATCAGGCGCAGGCCAACATTCAAATTCAGG 2942
DB 128 ----- 128
QY 2943 AAATATGAGAACACCACTAAAGATACCTCTCGAGAAACAATCCCAAGACACATAATCT 3002
DB 128 ----- 128
QY 3003 TCAGATTCCACCAAGTTGAAATGAAGAAAAATGTTAAGGCGCAGCCAGAGAGAAAGTT 3062
DB 128 ----- 128
QY 3063 GGGTTACCCACAAAGGGAAGCAATCAGACTAACAGCGGATCTCCCGCAGAAACCCCTAC 3122
DB 128 ----- 128
QY 3123 AAGCCAGAAGAGATGAGGGCCAATATTCACATCTTTAAGAAAAATATTTTCAACCCA 3182
DB 128 ----- 128
QY 3183 GAATTTTCATATCCAGCCAAACCAAGCTTCCTAAGTGAAGAGAGAAAAATAAATCTCTACAG 3242
DB 128 ----- 128
QY 3243 AGAAGCAAAATCTGACAGATTTTGTGCACCACCGGCTGCTTACAGAGCTCCTGAAG 3302
DB 128 ----- 128
QY 3303 GAAGCACCACATGGAAAGGAACAACTGGTACCAGCCACTGCARAAACATCCCAATTTGT 3362
DB 128 ----- 128
QY 3363 AAGACCAATTGATGCTATGAAGAAAGTGCATCACTAACGGGCAAAATAACAGCTAGTG 3422
DB 128 ----- 128
QY 3423 TCATAATGGCAGGATCAAAATTCACACATAATAATTAATTAACCTTAATTAATGGGCTAA 3482
DB 128 ----- 128
QY 3483 ATCCCAATTAAGACACAGACTGGCAAAATGGATAAGAGTCAAGACCCATCAGTGT 3542
DB 128 ----- 128
QY 3543 GCTGTATTTCAGGAGGCCATCTCAGATGAAAGACACACATAGGCTCAAAATAAAGGAT 3602
DB 129 -----LysAspThrHisArgLeuLysValLysGlyT 139
QY 3603 GGAGAGAGATTACCAAGTAATGGAACAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3662
DB 139 rparGlyIleTyrGlnGlyAsnGlyLysGln---LysLysAlaGlyValAlaIleLeu 158
QY 3663 TCTCTGATAAAACAGACTTTAAACCAACAAAGATCAAAAGAGACAAAGAGAGAGAGAGAGAG 3722
DB 158 alSerAspLysThrAspPheLysProThrLysIleLysArgAspLysGluGlyHisTyrM 178
QY 3723 TAATGGTAAAGGCATCAATGGAACAAAGAGAGTAACTATCTTAATATATACATGACCCCA 3782
DB 178 etMetValLysGlySerIleGlnGluGluLeuThrIleLeuAsnIleTyrAlaProA 198
QY 3783 ATACAGAGACCCAGATTCATTAAGCAAGTCTTAGAGACTTACAAAGAGACTTTGACT 3842
DB 198 snThrGlyAlaProArgPheIleLysGlnValLeuSerAspLeuGlnArgAspLeuAsp 218
QY 3843 CCCACACAATAATAGTGGGAGCTTAAATAATTAATAGACACTTTTAACACCCACCTGCCAA 3902
DB 127 ----- 128

Db 218 erHisThrLeuIleMetGly-----AspPheAsnThrProLeuSerT 232
QY 3903 TATTAGCCAGATCAATGACACAGAAATTAACAAGGATATCCAGGAGTTGAAGTACGCTC 3962
Db 232 hrLeuAspArgSerThrArgGlnLysValAsnAsnThrGlnGluLeuAsnSerAlaL 252
QY 3963 TGGACCAAGCGGACCTAATAGATATCTACAGACCTCCCAACCCCAATCAACAGAAATATA 4022
Db 252 euHisGlnAlaAspLeuIleAspIleTyrArgThrLeuTyrProLysSerThrGluTyrT 272
QY 4023 CACTCTCTCAGCATCACATATTCACCTATTTAAATATTCACCATGCTAAATTTAAAGTAAA 4082
Db 272 hrPhePheSerAlaProHisHisThrTyrSerLysIleAspHisIleLeuGlySerLysA 292
QY 4083 CACTCTCAGCAAAATGCAAAAGAACAGAAATCTTAACAACAGCTCTCTCAGACTACAGTG 4142
Db 292 laLeuLeuSerLysCysLysArgThrGluIleIleThrAsnTyrLeuSerAspHisSerA 312
QY 4143 CAATCTATTAGAACTCAGAAATTAAGAAACTCACTCAAAATCACAACTACATGGAAC 4202
Db 312 laMetLysLeuGluLeuArgIleLysAsnLeuThrGlnAsnCysSerThrTrpLysL 332
QY 4203 TGAACAACCTGCTCTCTGAATGACTACTGGTAAATAACAATAAGAGGCAAAATAAAGA 4262
Db 332 euAsnAsnLeuLeuAsnAspTyrGlyValGlnAsnLysMetLysAlaGluIleLysM 352
QY 4263 TGTCTCTTTGAAACCAATGAGAACACACAGACAAATGTACCAGATCTCTGGGCATATTTA 4322
Db 352 etPhePheGluThrAsnGluAsnLysAspThrTyrGlnAsnLeuTrpAspAlaPheL 372
QY 4323 AGCAGTGTGTAGAGGAAATTTATAGCACTAGATGCCCTACAGAGAAAGCAGGAAATAT 4382
Db 372 ysAlaValCysArgGlyLysPheIleAlaLeuAsnAlaHisLysArgLysGlnGluArgS 392
QY 4383 CTAAATATACACACTTAACATCACATTAAGAACTAGAGAGAGAGAGAGAGAGAGAGAG 4442
Db 392 erLysIleAspThrLeuThrSerGlnLeuLysGluLeuGluLysGlnGluInThrHis 412
QY 4443 CAAAAGCTAGCAG 4502
Db 412 erLysAlaSerArgSerGlnGluIleThrLysIleArgAlaGluLeuLysGluIleGlnT 432
QY 4503 CACAAAAGCCCTTCAATTAATCAATGAATCCAGGAGCTGGTGTGTTTCAAAAGATCAGC 4562
Db 432 hrGlnLysThr-PheGlnLysIleAsnGluSerArgSerTrpPheGluArgIleAsn 451
QY 4563 AAAAT--AGACCCTAGACACTAATAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 4618
Db 452 LysIleAspArgProLeuAlaArgLeuIleLysLysArgGluLysAsnGlnIleAsp 471
QY 4619 GCAATAAAATGATAAGGGGATATCACCCGATCCACAGAAATACAAACTATTATC 4678
Db 472 AlaIleLysAsnAspLysGlyAspIleThrThrAspProThrGluIleGlnThrIle 491
QY 4679 AGAGAAATATTAACACCTCTATGCAATAAATACTAGAAATCTAGAGAAATGGATAAA 4738
Db 492 ArgGluTyrTyrLysHisLeuTyrAlaAsnLysLeuGluAsnLeuGluMetAspThr 511
QY 4739 TTCCTGGACACATATGACCTGTATGGACCTTGGGGGACAGACAGAGAGGGGGAATG 4798
Db 512 PheLeuAspThr----- 515
QY 4799 CAGAAATAAAGACAAAGACAAAGAGATGTTGGAAGTAGGGGTGAGGGGGCAACTTG 4858
Db 515 ----- 515
QY 4859 CCTCTAATGGACAAGGCCCTGAGCTTTACACCACCTCTGTTATTATTAGCAAAAGAG 4918
Db 515 ----- 515
QY 4919 ATAGCGAGGGGTGAGTTGGAAGAGAGGTACGCTGTTAGTCCAGAGTAGCCCTGCAAG 4978
Db 515 ----- 515

QY 4979 ACTGCATTCTCAACAATAGGCTCTAGATGTCCCAAGTAGATAAACCCTCAAGAGCCAGTG 5038
Db 515 ----- 515
QY 5039 CCAGGAGTGATGGCCCTCAGCAAACTTCTTAGCGCAGGCACAGAGAAGTAAGTTTGGCCAC 5098
Db 515 ----- 515
QY 5099 ATTCTGTATTACAGATAAACAAGTTTGCTGTTGATCAAGTAGCTCCAGTGGGAATCTGA 5158
Db 516 -----Cys--- 516
QY 5159 GTTGCTCATGCTCTTTTGGCCCTTTTGGCTCCCAAAACACATACACCTCTCAAGACTA 5218
Db 517 -----ThrLeuProAlaGluLeu 521
QY 5219 AACCCAGGAAGAAGTCAAAATCCCTGATATATACAGTAACAAGTTCTAAAATTGAGACAGTA 5278
Db 522 AsnGlnGluGluValGluSerProAsnArgProIleThrGlySerGluIleValAlaIle 541
QY 5279 ATTGATAGCTTACCAACCAAAAAAGTCCAGGACACAGCGGATTCACAGCCAAATTTCTAC 5338
Db 542 IleAsnSerLeuProThrLysLysSerProGlyProAsnGlyPheThrAlaGluPheTyr 561
QY 5339 CAGAGTCAACAAGAGAGAGCTGCTACTTCTTCTGAAACTATTCCAAAAAATAGAA--- 5395
Db 562 GlnArgTyrLysGluGluLeuValProPheLeuLeuLysLeuPheGlnSerIleGluLys 581
QY 5396 AATGGGAATCTCCCTTAATCAATTTACAGGCGCAGCATCTCTGATACCAAAACCTTAG 5455
Db 582 Glu-GlyIleLeuProAsnSerPheTyrGluAlaSerIleThrLeuIleProLysAlaG 601
QY 5456 CAGTGAACAACAAGAGAGAAATTTCAAGGCCATATCCCTGATCAACATTCATGCTGAA 5515
Db 601 yArgAspThrThrLysLysGluAsnPheArgProIleSerLeuMetAsnIleAspAla 621
QY 5516 AATCTCTCAATAAATACTGGCAACCAATCCAGCAGCAGCATCAAAAAGCTTATCTACCA 5575
Db 621 sIleLeuAsnLysIleLeuAlaHisArgValGlnGlnHisIleLysLysLeuIleHisI 641
QY 5576 TGATCAAGTTGGCTCATCCCTGGGATCGGATCGAAGCTGGTTCAAAATATGCAAAATCAA 5635
Db 641 sAspGlnValGlyPheIleProGlyMetGlnCysTrpPheAsnIleCysLysSerIleAs 661
QY 5636 TGTAGCCCATCACATAAACAGAACCAATCAGAAAAACCATGATTATCTCAATAGATGC 5695
Db 661 nValIleGlnHisIleAsnArgThrLysAspLysAsnHisMetIleIleSerIleAspAl 681
QY 5696 AGAAAAGGCTTTGTGCTCAAAATTCACAGCCCTTCATGCTTAAATAATTTCTCAGTAAACTAGG 5755
Db 681 aGluLysAlaPheAspLysIleGlnProPheIleLeuLysThrLeuAsnLysLeuG 701
QY 5756 TATCGATGAATGTATCTCAAAATAATAAGAGCTATTATTAC-AAACCCACAGCCCAATAT 5814
Db 701 yIleHisGlyThrThrLeuLysIleIleArgAlaThrTyrAspLysProThrAspAsnI 721
QY 5815 CATACTGAATGGCAAAAACTGGAACCTTCCTTTGAGAAGCTGGCAGACAGACAGAGATG 5874
Db 721 elysLeuAsnGlyGlnLysLeuGluAlaPheProLeuLysThrAspThrArgGlnGlyCy 741
QY 5875 CCCTCTCTCACCACCTCTATTCAAGATACTATTGGAAGTTCTGGCCAGGCGCAATCAGGCA 5934
Db 741 sProLeuSerProLeuLeuPheAsnIleValLeuGluValLeuAlaArgAlaIleArgG 761
QY 5935 ATAGAAGAAATAAAGGGTATTCAATAGAAAGAGAGAGAGATTTGCTCTCTGTTTGC 5994
Db 761 nGluLysGluIleLysGlyIleGlnLeuGlyLysGluGluValLysLeuSerLeuPheAl 781
QY 5995 AGATGACATGTTGTATATTATAGAAACCCCTCGTCTCAGGCCCAAAACTCCTTAAGCT 6054
Db 781 aAspAspMetIleValTyrLeuGluAsnProIleValSerAlaGlnAsnLeuLeuLysLe 801

QY 6055 GATAAGCAACTTCACGAAGTCTCAGGACACAAAATCAATGTGCAAAAATCACAGCATT 6114
Db |||||||
QY 801 uileSerAsnPheserLysValSerGlyTyrlLysileAsnValGlnLysSerGlnAlaPh 821
QY 6115 CTTATACCCCAATATAGACAAAACAGAGAGCAAAATCATGAGTGAACCTCTCATTCACAAAT 6174
Db |||||||
QY 821 eLeuTyThrAsnLysArgGlnThrGlnSerGlnileMetSerGluLeuProPheThrIl 841
QY 6175 TGCTCAAAAGAGATAAAATACCTAGGAATACAACTTACAAGGACAGCTAGGAACTCTT 6234
Db |||||||
QY 841 eAlaserLysArgGlyTyrlLeuGlyileGlnLeuThrArgaspValLysaspLeuPh 861
QY 6235 CAAGGAGAACTCAAAACCACTGATCAAGGAAATAAGAGAGACACAAACAAATGGAAAA 6294
Db |||||||
QY 861 eLysGluasnTyrlLysProLeuLeuLysGluileLysGluaspThrAsnLysTrpLysAs 881
QY 6295 CATTCATGCTCACAGATAGTAAGATCAT-----GAAATGCCATAGTCCCAAGT 6347
Db |||||||
QY 881 nileProCysSerValGlyArgileSerPheValLysMetAlaileLeuProLysVa 901
QY 6348 AAATATAGATTGCTGCTACCCCACTCAAGCTACCATTTGCTTCTTCACAGAAATGGA 6407
Db |||||||
QY 901 lileTyArgPheAsnAlaileProLileLysLeuProMetPhePheProGluLeuGl 921
QY 6408 AAAACAACTTTAAATTCATATGGAACCAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6467
Db |||||||
QY 921 uLysThrThrLeuLysPheileTrpAsnGln-LysArgAlaArgileAlaLysSerileL 941
QY 6468 TAAGCAAAAGAAAGAAAGCTGGAGTATCATGCTACCTGCTTAACTTAACTATATAGG 6527
Db |||||||
QY 941 euSerGlnLysAsnLysAlaGlyGlyileThrLeuProaspPheLysLeuTyrrLysA 961
QY 6528 CTACAGTAACCAAACTGCATGGTACTGGTACCAAAAACAGATATAGACCAATGGGAACA 6587
Db |||||||
QY 961 laThrValThrLysThrAlaTrpTyrrTyrlGlnAsnArgaspLeaspGlnTrpAsnA 981
QY 6588 GAACAGAGACCTCAGAAATACACT-GCAATCTACATCCATCTGATCTTTGACAAACCTG 6646
Db |||||||
QY 981 rgThrGluProSerGluileThrProHileTyrrAsnTyrlLeuilePheAspLysProG 1001
QY 6647 ACAAACAAGCAATGGAAGAAGGATCCCTATTATTAATGATGTTGGAAGAAGCTGGC 6706
Db |||||||
QY 1001 luLysAsnLysGlnTrpGlyLysaspSerLeuPheAsnLysTrpCysTrpGluAsnTrpL 1021
QY 6707 TAGCCATATGCAAGAAGCTGAACTGGATCCCTTCTTACACCTTATACAAAAGTTAACT 6766
Db |||||||
QY 1021 euAlaileCysArgLysLeuLysLeuaspProPheLeuThrProTyrrThrLysileAsn 1041
QY 6767 CAAGATGAATTAAGACTTAATATAGACATAAAACCAATAAAACCCA-GAAGAAAACC 6825
Db |||||||
QY 1041 erArgTrpIleLysAspLeuAsnValArgProLysThrIleLysThrLeuGluGluAsnL 1061
QY 6826 TAGCAATACCATTCAGGATATGACATGGCAAGACTTCATGACTTAAACACCAAAAG 6885
Db |||||||
QY 1061 euGlyileThrileGlnaspIleGlyMetGlyLysaspPheMetSerLysThrGlnLysA 1081
QY 6886 CAATGGCAACAAAACCCCAATAGACAAGTGGGATCTGATTAAACTATAGACTTCTGCA 6945
Db |||||||
QY 1081 laMetAlaThrLysAlaThrileaspLysTrpaspLeuileLysLeuArgSerPheCyst 1101
QY 6946 CAGCAAAAAAACTGCTCATCAGAGTGAACAGCAACCTACAGAAATGGGAGAAAATTTT 7005
Db |||||||
QY 1101 hrAla-LysGluThrThrIleArgValAsnArgGlnProThrLysTrpGluLysilePhe 1120
QY 7006 GCAATCTATCGATCTGACAAAGGCTAATATCCAGAGATCTACGAGAACTTAAACAAAT 7065
Db |||||||
QY 1121 vallleTyrrSerAspLysGlyLeuilePheArgileTyrrAsnGluLeuLysGlnile 1140
QY 7066 TACAAGAAAAA-----AACAACCCCGTCAAAATATGGCAAGGATATGACAGACACTTC 7121
Db |||||||
QY 1141 TyrlLysLysLysThrAsnAsnSerileLysLysArgAlaLysaspMetAsnArgHisPhe 1160
QY 7122 TCAAAAGAGACATTTATGACGCCCAACAACATATGAAAAAACCTCATCATCTGTC 7181

Db ||||||| SerLysGluaspIleTyrrAlaAlaLysAsnHisMetLysLysCysSerSerSerLeuAla 1180
QY 7182 GTTAGAGAAATGCAAAACCAACACACTGACATACCATCTCTAGTGTAGTAGAGTGTG 7241
Db |||||||
QY 1181 ileArgGluMetGlnileLysLeuThrMetArgTyrrHisLeuThrProValArgMetAla 1200
QY 7242 ATCACTAAAAGTGCAGGAAACAAATGCTGGAGAGGATGTGGAGAAATAGGACACTT 7301
Db |||||||
QY 1201 ileileLysLysSerGlyAsnAsnArgCysTrpArgGlyCysGlyGluileGlyThrLeu 1220
QY 7302 TTCACCTGTTGGTGGGAATGTAAATAGTTCAACCATTTGTTGGAAGACAGCTGGAGATTC 7361
Db |||||||
QY 1221 LeuHisSerTrpTrpAspCysLysLeuValGlnProLeuTrpLysSerValTrpArgPhe 1240
QY 7362 CTTAAGGATCTAGAACCAAGAAATATCTATTATTAAGACACATGCCCAATCTACTGAGTATATAC 7421
Db |||||||
QY 1241 LeuArgLysLeuGluLeuProPheaspProAlaileProLeuArgGlyileTyrr 1260
QY 7422 CCAAGGAATATAATCATCTATTATAAGACACATGCACACATATGTTTATTGCGACA 7481
Db |||||||
QY 1261 LeuLysaspTyrrLysSerCysTyrrLysaspThrCysThrCysMetPheileAla 1280
QY 7482 CTGATCACATAGCAAAAGACTTGGACCAACCCCAATGTCCATCAGTATAGTGGATA 7541
Db |||||||
QY 1281 LeuPheThrileAlaLysThrTrpAsnGlnAlaLysCysSerThrMetileaspTrpIle 1300
QY 7542 AAGAAAACATGGCCACATATACCATGAATATCTATGCAGCATATAAAG-GATGAGTTC 7600
Db |||||||
QY 1301 LysLysMetTrpHisileTyrrThrMetGluTyrrAlaAlaileLysAsnaspGluPhe 1320
QY 7601 ATGTCCTTTGACAGATATGATGAAGCTGGAAACCATCATTTCTACGAACTAACACAA 7660
Db |||||||
QY 1321 MetSerPheValGlyThrTrpMetLysLeuGluileLysLeuSerLysLeuSerGln 1340
QY 7661 GAACAGAAAACCAACACCATGTTCTCCTACTTGTAGTGGGATTTGACAAATGAGAAGA 7720
Db |||||||
QY 1341 GluGlnLysThrLysHisArgilePheSerLeuileGlyGly----- 1354
QY 7721 CATGGACACAGGAGGGG-----AACATCACACACAGGCTCTGTTT 7762
Db |||||||
QY 1355 Asn***HisLysThrGlyMetProSerLeuThrThrProLleGlnHisSerValGlySer 1374
QY 7763 GTGGGTGCGGACTAGGAGAGGATAGCATATTAGAGAAATACCTAACTAGTACGCGGT 7822
Db |||||||
QY 1375 SerGlyGlnGlyAsn***AlaGlyGluGlyAsnLysGlyTyrrSerileArgLysArgGly 1394
QY 7823 TGATGGGTGCAGCAAGCCACCATGGCACATGTATACCTATGCTAACAACCT 7873
Db |||||||
QY 1395 SerGlnileValProValCysArg***HisaspCysileSerArgLysPro 1411
RESULT 12
ABG03983
ID ABG03983 standard; Protein; 1691 AA.
AC ABG03983;
DT 13-FEB-2002 (first entry)
DE Novel human diagnostic protein #3974.
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
OS Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

DR N-PSDB; AAS68170.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity

XX Claim 20; SEQ ID No 34342; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1691 AA;

Alignment Scores:

Pred. No.: 0 Length: 1691
Score: 4893.50 Matches: 1124
Percent Similarity: 60.58% Conservatives: 147
Best Local Similarity: 53.57% Mismatches: 313
Query Match: 9.35% Indels: 522
DB: 22 Gaps: 32

US-10-083-853B-2 (1-29921) x ABG03983 (1-1691)

QY 1210 CATTCCCACTGAGTACCCAGTTCATCTC----- 1239
DB 8 HisProProAlaGluThrProArgHisValLeuLeuLeuMetAlaAspSerLysAsnGlu 27
QY 1240 -----ACTGGACTGGTTAGACATGGGTGCAGCCACCAGGAAGGTGAGCTGAA--- 1287
DB 28 LeuGlnAsnMetAsnThrGlyGlnThrCysGlyThrArgGlnThrLeuLysSerLysLeu 47
QY 1288 GCAGGTGGGTCTCCCTCAGCCGCGAGTGCAGGGGTGGGGATCTCCTTCCCCCA 1347
DB 48 GlnGlyTrpAsnLysPro-----GlyLeGlnSerAspPro 59
QY 1348 GCCAAGGGAAGCATGAGACACTGATACAGAGGAATGGTGCACTCTAGTCCAGACTG 1407
DB 60 GlySerGlnGlyHisProArg-CysLeuLys----- 69
QY 1408 CACTTTCCCATAGTCTTTGCAACTGGCAGACACAGGAGATTTCCCCAGTGCCTATGCCA 1467
DB 70 -ThrPheProLeuValAlaGlnSerSer-----ThrGlnGluArgAsnHi 84
QY 1468 C-----CAGGGCCCTGGGTTTCAACACAAACTGGGGCCCATTT----- 1507
DB 84 sValMetGluLysAlaValLeuLysSerGlnAsnTrpAlaIleGluPheIleTh 104

QY 1508 -----TGGACAGACACCGAGCTAGCCGACGACAGTTTATTTTTCATACCCAGTGG 1557
DB 104 rLysIleGlyTrpPheProAspTyrSerGlyLysValLysTyrLeuSerSerProLysSe 124
QY 1558 CGCCTGGAAATGCCAGCAGACAGAACCATTCACCTCCAGGATCAAGTGTCTGGCTCAG 1617
DB 124 rSer-----CysThrThrValLeuGlnValThrHisPheSe 136
QY 1618 T---GGGTCCACCCCATCGAGCCCGAGCTA---GCTAAGATCCAC----- 1657
DB 136 rLeuPhePheHisProAsnProAspGluLeuValSerLysIleLysValTrpSerLysHi 156
QY 1658 -----TGGCTTGAATTCCTCTCCGACGACGACAGCAGTC 1689
DB 156 sArgLeuTyrGlnAsnAsnSerGluAlaPheLeuGluValGlnIleProGluProLysCy 176
QY 1690 TGAGATTGA-----CTGGG---ATGCTTGAGCTTGG 1718
DB 176 sGluValTyrSerMetArgThrMetGlyArgArgGlnProGlyThrAlaMetAspLeuAs 196
QY 1719 TGAGGGGAGGGCGCTGCTGCTTGCCTTGCCTGAGCTTGCAGTAGCGGCGGTTCACCTCAA 1778
DB 196 nAlaGlyGlySerIleCysAsnVal----- 204
QY 1779 AGTGTAAACAAGACTACTGGGAAGTTTGAATGGGCGCCACCGCAGCTCAGCAAGCCG 1838
DB 205 -----ProArgThrLeuLeuHisLeuThrGlyLeuSerThrPheArgAspArgGlnAr 222
QY 1839 CTGTGGCAAACTGCTCTCTAGATTCTCTCTTTTGGGCGAGGTCTCTCTGAAGAAGG 1898
DB 222 gValGly-----AlaTyrLeuAspIleArgG1 231
QY 1899 CACAGCCCCAGTCAGGGACTTATAGATAAAACCCCATCTCCCTG---GGACAGACA-- 1953
DB 231 yTrpTyrIleLeuValLeuValValAspLysGluLysLeuGlyIleProGlyGlnLysG1 251
QY 1954 -----ACCTGGGGG-----AAGGGTGGCTGTGGGTGCAGCT 1985
DB 251 nAsnSerSerLysIleLeuPheGlySerGluLysLeuLysGlnTrpValTyrValLys-- 270
QY 1986 TCTCAGACTTAAACATTCCTGCTGCTGAGGCTCTGAAGAGAGACGCGGATCTCCAGACA 2045
DB 271 -----As 271
QY 2046 GCATTTTCAGCTCTGATAAGGAGCAGGCTGCTCCTCAAGTGGGTCCCTGACCCCATGTA 2105
DB 271 nHisAlaLysGluSerIleThrGlnGluAlaMetGly----- 283
QY 2106 TCCTGACTGGGAGACATCTCCCATTAGGGGCCAATAGACATTTTCATACAGGAGACAGGT 2165
DB 284 -----I1 284
QY 2166 CTGAGTGGACCTCCAGCAAACTCCAGCAGACCTGCAGCAGAGCGCGCTGACTGTAGAA 2225
DB 284 eTrpValGlyGln----- 288
QY 2226 GGAAGAGTAACAACAGAAAGGAATAGTATCAAC-----ATTACAA 2267
DB 289 -----GlnArgLysArgGlnProLeuGlyTyrGluGluArgLysLeuThrAsnAr 305
QY 2268 AAAGGACATCCACTCAGACACCCCATCTGAAGGTCAACACACAAAGACCAAGGTAA 2327
DB 305 gLysAspIleHisIleLysAsnProSerValCysHisHisGlnArgProLysValas 325
QY 2328 TAAACCAAAAGATGGGAAAAC-CAGTGCAGAAACACTGAAATTTCCAAAACCCAGAA 2386
DB 325 pLysThrThrLysMetGlyLysLysGlnSerArgLysThrGlyAsnSerLysAsnGlnSe 345
QY 2387 CTCTCTTCTCAACCAAGGATCACAACCTCTCCGACGAGGAGGACAAACACCATGGA 2446
DB 345 rThrSerProThrLeuLysGluSerSerSerSerProAlaThrGluGlnSerTrpThrG1 365

QY	2447	GAATGAGTTTGAGGAATTGACAGAAAGTAGGGCTTCAGAGGTGGGTAATAACAAACTCTC	2506		Db	593	-----AspThrHisar	596
Db	365	uAsnAspPheAspGluLeuArgGluGluGlyPheArg	382		QY	3586	GCTCAAAATAAAGGATGGAGGAAGATTACCAAGTAAATGGAAGAAAAAAGAAC	3645
QY	2507	CGAGCTAAAGGACGATGTTCTAACCCATGCAAGCAAGCACTAAGAACCCTTGAAAAAGGTT	2566		Db	596	gLeuLysIleLysGlyTrpArgLysIleTrpGlnAlaAsnGlyLysGln---	596
Db	382	rgLLeuGlnGluGluValGlnAsnAsnGlyLysGluValLysAsnPheGlyLysLysLe	402		QY	3646	AGGGTTGCAATCCCTAGTCTCTGTATAAACAGACAGCTTTAAACCAACAAAGATCAAAAGAGA	3705
QY	2567	AGATCAATTTGCTAACTAGATAATCAGTGTAGAGAGAACATTAATAGACCTGATCGAGCT	2626		Db	615	aglyValAlaIleLeuValSerAspLysThrGluPheLysProThrLysIleLysArgAs	635
Db	402	uasplurprileThrArgIleThrAsnAlaLysLysSerLeuLysAspLeuMeGluLe	422		QY	3706	CAAGAAGCCATTACATAATGCTAAAGGCATCATGGAAGCAAGAGAGCTTAACCTCT	3765
QY	2627	GAAGAACCAAGACAGAACTTCATGAAGCATACACAAGCTTCATAGCCAAATCGATCA	2686		Db	635	plysGluGlyHisTrpIleMetValLysGlySerIleGlnGlnGluLeuThrIleLe	655
Db	422	uLysThrLysAlaArgGluLeuCysAspGluArgThrSerLeuSerSerGlnCysAsnG	442		QY	3766	AAATATACATGACCCCAATACAGGACGCCAGATTTCATAAAGCAAGTTCCTTAGAGACCT	3825
QY	2687	ACCAGAAAGAGATATCAGTCATTCAGATCAATTAATAAAGAAAGTGAAGACAA	2746		Db	655	uasnlleTrpAlaProAsnThrGlyAlaProArgPheThrLysGlnValLeuArgAspLe	675
Db	442	nLeuGluGluArgValSerValMetGluAspGluMetAsnGluIleLysGlnGluGlu	462		QY	3826	ACAAGAGACTTTGACTCCACACAATAATAGTGGAGTCTAAATAATAATAGACACTT	3885
QY	2747	GATTACAGAAAAAGAGTGAAGAAACAAAGAGCCCTCCAAAGATTATGGGACTATCT	2806		Db	675	uGlnArgAspLeuAspSerAsnThrIleIleThrGly-----AspPh	689
Db	462	sPheArgGluLysArgIleLysArgAsnGluGlnSerLeuGlnGluIleTrpAspTyMe	482		QY	3886	TAACACCCACTGCCCAATATTAGGCAGATCAATGAGACAGAAATTAACAAGATATCCA	3945
QY	2807	GAAGAACCAATCTACATTTGATTGGTGTGCTCCCAAGATGATGGGAGATGGATCAA	2866		Db	689	easnlThrProLeuSerThrLeuAspArgSerMetArgGlnLysValAsnLysAspIleG	709
Db	482	tlYsArgProAsnLeuArgLeuIleGlyValProGluSerAspGlyGluAsnGlyThrly	502		QY	3946	GGAGTTGAACTGAGCTCTGGACCAAGCGGACCTAATAGATATCTACAGAACTCCCAACC	4005
QY	2867	GTGGAACACACTCTTCAGGTATTATCCAGGAGAAATTCGCCAT-CTATCAGGCGAGC	2925		Db	709	nGluLeuAsnSerAlaLeuHisGlnAlaAspLeuIleAspIleTrpArgThrLeuHisPr	729
Db	502	sLeuGluAsnThrLeuGlnAspIleIleGlnGluAsnPheProAsnLeuAlaArgGlnAl	522		QY	4006	CAATCAACAGAAATATACACTCTCTCAGCATCATATACACTATTTTAAATATGACCA	4065
QY	2926	CAACATTCAAATTCAGGAAATATGGAACACCACTAAAGATACTCTCGAGAAACAAT	2985		Db	729	olYsSerThrGluAsnThrPhePheSerAlaProHisIleThrTrpSerLysIleAspHi	749
Db	522	aAsnIleGlnIleGlnGlnArgThrProGlnArgTySerLeuArgArgAlaTh	542		QY	4066	TCTAATTTTAACTAAACACTCTCTCAGCAATTCGCAAAAGAACAGAACTCTAACAAACAG	4125
QY	2986	CCCAAGACACATAATCTCAGATTCACCAAGGTTGAAATGAAGAAAAATTTAAGGC	3045		Db	749	sIleValGlySerLysAlaLeuLeuSerLysCysArgLysGluIleIleThrAsnCy	769
Db	542	rProArgHisIleIleValArgPheThrLysValGluMetLysGluLysMetLeuArgVa	562		QY	4126	TCTCTCAGACTACAGTGCATCTATTAGAATCAGAATTAAGAACTCCTCAAAATCA	4185
QY	3046	AGCCAGAGAAAGGTTGGTTTACCACAAAGGGAAGCAATCAGACTAACAGCGATCT	3105		Db	769	sLeuSerAspHisSerAlaIleLysLeuGluLeuArgIleLysLysLeuProGlnAsnCy	789
Db	562	lAlaArgGluLysGlyArgValThrHisLysGlyLysProValArgLeuThrAlaAspLe	582		QY	4186	CACACTACATGGAACAGAACACTGCTCTCTGATGACTACTGGTAAATAACAATAAT	4245
QY	3106	CCCGCAGAAACCCCTACAGCCAGAGAGAGTGGGCGCAATATTCCACATTTCTTAAAGA	3165		Db	789	sSerThrIleTrpLysLeuAsnAsnLeuLeuAsnAspTyTrpValHisAsnGluMe	809
Db	582	uLeuAlaGluThrLeuGlnAlaArgArgGln	592		QY	4246	GAAGCAAAATAAAGATGTTCTTTCAACCAATGAGAACAAAGACACAAATGTACCAGAA	4305
QY	3166	AAATAATTTCAACCCAGAAATTCATATCCAGCCCAACCAAGCTTCCTTAAGTGAAGAGA	3225		Db	809	tlYsAlaGluIleLysMetLeuPheGluThrAsnGluAsnLysAspThrThrTyTrpGlnAs	829
Db	592	-----	592		QY	4306	TCTCTGGGCGCATTTTAAAGCAGTGTGTAGAGGAAATTTATAGCACTAGATGCCCTACAA	4365
QY	3226	ATAATAATCTCTACAGAGAGCAAAATGCTGACAGATTTTGTACCAGCGCTGCT	3285		Db	829	nLeuTrpAspThrLeuLysAlaValCysArgGlyLysPheIleAlaLeuAsnAlaHisG	849
Db	592	-----	592		QY	4366	GAGAAGCAGGAAATATCTAAATAAGACACCTTACATCATCACAATTAAGAAGACTAGAGA	4425
QY	3286	TACAAGAGCTCCTGAAGGAAGCAACATGGAAGGAACAACTGGTACCAGCCACTGCA	3345		Db	849	YargLysGlnGluArgSerLysIleAspThrLeuThrSerGlnLeuLysGluLeuAlaLy	869
Db	592	-----	592		QY	4426	GAAGAGCAAAACAAATTCAAAAGCTAGCAGAGAAAGAAATAACTTAAGTACAGAGCAGA	4485
QY	3346	AAACATCCCAATTTGTAAGACCAATTGATGCTATGAAGAAGTGCATCACTAACGGGC	3405		Db	869	sGlnGluGlnThrHisSerLysAlaSerArgArgGlnGluIleThrLysIleArgAlaG	889
Db	592	-----	592		QY	4486	ACTGAAGGAGATAGACACAAAAAGCCCTTCAAAATAAATCAATCAATCCAGGAGCTGT	4545
QY	3406	AAATAACCAAGCTAGTGTATTAATGGCAGGATCAAAATTCACACATAAATAATTAACCTT	3465		Db	889	uLeuLysGluIleGluThrGlnLysThrLeuGlnLys-IleAsnGluSerArgSerTrp	909
Db	592	-----	592		QY	4546	TTTTTGAAGATCATCAGCAAAATAGACACTAGACAGACTAATAAGAAGAAAGAGAGAA	4605
QY	3466	AAATGTAATGGCTAAATTCGCCAATTAAGACACACAGACTGGCAAAATTTGGATAAGAG	3525		Db	909	he------Leu--ProLeuAlaArgLeuIleLysLysLysThrGluLy	922
Db	592	-----	592		QY	4606	GAATCAAGAGATGCAATAAATAAAGGATATCACCAGCTCCACAGAGAAAT	4665
QY	3526	TCAAGACCCATCAGTGTCTGTATTTCAGAGGCCCATCTCACATGAAAAAGACACACATAG	3585					

Db 922 sAsnGlnIleAspAlaIleLysAsnAspLysGlyAspIleThrThrAsnProThrAspIle 942
QY 4666 ACAAACTATTATCAGAGAAATATTATAAACACCTCTATGCAATAATAAATAGAAAATCTAGA 4725
Db 942 eGlnThrAlaIleArgGluTyrThrLysHisLeuTyrAlaAsnLysLeuGluAsnLeuGI 962
QY 4726 AGAAATGGATAAATTCCTGACACATATGTAGCCTGTATGGACCTTGGGGGACAGAACAA 4785
Db 962 uGluMetAspLysLeuLeu----- 968
QY 4786 AAGGGGTGAATCAGAGAAATAAAGACAAAGACAAAGAGATGTTTGGNAGTAGGGTC 4845
Db 968 ----- 968
QY 4846 AGGGGGCAACTTGCCTCTAATGGACAAGGGCCCTGAGCTTTACACCACCTCTGTATTTA 4905
Db 968 ----- 968
QY 4906 TTAGGCAAAGAGATAGCGAGAGGGTGAGTTGGAAGAAGAGGTGCTGTTAGGTCCAGA 4965
Db 968 ----- 968
QY 4966 GTAGGCCTCAAGACTGCATTCCTCAACAATAGGCTCTAGATGTCCAGTAGATAACCT 5025
Db 968 ----- 968
QY 5026 CAAGGAGCCAGTGCCAGGAGGTGATGGCCCTCAGCAAACTTCTAGGGCAGGCACAGAAG 5085
Db 968 ----- 968
QY 5086 TAAGTTGCCCATCTCTGTATTTCAGATAAACAGTTGCTGTTGATCAAGTAGCCTCC 5145
Db 968 ----- 968
QY 5146 AGTGAATGCTGAGTTGGTCATGATCCCTTTGGCCCTTTTGGCTCCCAAAACACATACAC 5205
Db 969 -----AsnThrTyrThr 972
QY 5206 CCTCTCAAGACTAAACAGGAGAGAGTCAAAATCCCTGAATATACCAAGTAACTCTAA 5265
Db 972 rIleProArgLeuAsnGlnGluValGluProArgAsnArgProIleThrGlySerGI 992
QY 5266 AATTGAAGCAGTAATTGATAGCTTACCACCAACCAAAAGTCCAGGACGAGCGGATTCAC 5325
Db 992 uIleGluAlaIleAsnArgLeuProThrLysLysCysProGlyProAspGlyPheTh 1012
QY 5326 AGCCAATCTACACAGGTACAAAGAGAAGCTGTACTATTCTCTCTGAAACTATTCCA 5385
Db 1012 rAlaGluPheTyrGlnArgTyrLysGln-----ValProPheLeuLysLeuPheGI 1030
QY 5386 AAAATAGAA---AATGGGAATCCTCCCTAACTCATTTTACGAGGCCAGCATCATCTGA 5442
Db 1030 nSerThrGluLysGly-GlyIleLeuProLysSerPheTyrGluAlaSerIleIleLeuI 1050
QY 5443 TACCAAACTAGCAGTGCACACACAAAGAGGAATTTACGGCCCATATCCCTGATGA 5502
Db 1050 leProLysProGlyLysAspThrThrLysLysGluAsnPheArgProIleSerLeuMeta 1070
QY 5503 ACATTGATGTGAAATCCCTCAATAAATACTGGCAACCAACCAATCCAGCAGCACATCAAAA 5562
Db 1070 snIleAspValLysIleLeuAsnLysIleLeuAlaLysGlnIleGlnHisIleLysL 1090
QY 5563 AGCTTATCCACATGATCAAGTTGGCGTCATCCCTGGGATGCAAGCTGGTTCAAAATAT 5622
Db 1090 ysLeuIleHisAspGlnValGlyPheIleProGlyMetGlnArgTrpPheAsnIleA 1110
QY 5623 GCAATCAATAATGTAGGCCATCACAATAACAGAACCAATGACAAAAACACCATGATTA 5682
Db 1110 rGlySerIleAsnValIleGlnHisIleAsnArgThrLysAspLysAsnArgMetIleI 1130
QY 5683 TCTCAATAGTCAGAAAAGGCTTTGTCAAAATTCACAGCCCTTCATGCTTCAAAATTC 5742
Db 1130 leSerIleAspAlaGluLysAlaPheAspLysIleGlnArgPheMetLeuLysThrL 1150

QY 5743 TCAGTAAACTAGGTATCGAATGTATCTCAAAATAATAAGAGCTATTATTATAC-AAAC 5801
Db 1150 euAsnLysLeuGlyIleAspGlyThrTyrLeuLysThrIleArgAlaIleTyrAspLysP 1170
QY 5802 CCACAGCCCAATATCATACTACTGGAATGGGCAAAAACATGGAGCATTCCTCTGAGAACCTGGCA 5861
Db 1170 roThrAlaAsnIleIleLeuAsnGlyGlnLysLeuGluAlaPheProLeuLysThrGlyT 1190
QY 5862 CAAGACAAGATGCCCTCTCTACCACTCTTCAAGTACTATTATGGAGTTCGTGCCA 5921
Db 1190 hrArgGlnGlyCysProLeuSerProLeuLeuPheAsnMetValLeuLysLeuAla 1210
QY 5922 GGCAATCAGGCAATAGAAAGATTAAGGGTATTCAAAATAGAAAGAGAGGAAGTCAAT 5981
Db 1210 rgAlaIleSerGlnGluLysGluIleLysaspIleGlnLeuGlyLysGluGluValLysL 1230
QY 5982 TGTCTCTGTTGGCAGATGACATGTTGTATATTTAGAAAACCCCTCTCTCAGGCCAAA 6041
Db 1230 euSerLeuPheAlaAspAspMetIleValTyrLeuGluAsnProIleValSerAlaGln 1250
QY 6042 AACTCCTTAAGCTGATAAGCACTTCAGCAAGTCTCAGCACACAAAATCAATGTCAAA 6101
Db 1250 snArgLeuLysLeuIleSerAsnPheArgLysValSerGlyTyrLysIleAsnValGlnL 1270
QY 6102 AATCACAGCATTTCTTATACGCCAATAATAGACAAACAGAGAGCCAAAATCATGAGTGAAC 6161
Db 1270 ysSerGlnAlaPheLeuTyrThrAsnAsnSerGlnThrGluSerGlnIleMetSerGluL 1290
QY 6162 TCTCATTCACATTTGCTACAAAGAGATAAATACCTAGATATACAACTTACAGGAC 6221
Db 1290 euProPheThrAlaSerLysArgIleLysTyrLeuGlyIleGlnLeuThrArgAsp 1310
QY 6222 CGTAGAATCTTCAAGGAGAGACTACAAACCTGATCAAGGAAATAAGAGAGAGACACAA 6281
Db 1310 alLysAspLeuPheLysGluAsnTyrLysProLeuLeuSerGluIleLysGluAspThrA 1330
QY 6282 ACAAAATGGAATAACATTCCTCATGCTCACAGATA-----GTAAGAATCATGAAAATGCC-A 6334
Db 1330 snLysCysLysAsnValProCysSerTrpValGlyArgIleAsnIleMetLysMetAlaI 1350
QY 6335 TACTGCCCAAGTAATTAATAGATTTCAGTGCCTACCCCTCAGCTTACCATTCACATTCCT 6394
Db 1350 leLeuProLysValIleTyrArgPheAsnAlaIleProIleLysLeuProMetThrPheP 1370
QY 6395 TCACAAATGGAATAACAACTTTAAATTTTCATATGGAACCAACCAAGAGAGAGAGAGAG 6454
Db 1370 heThrGluLeuGluLysThrThrLeuLysPheIleTrpAsnGln-LysArgAlaCysIle 1389
QY 6455 GCCAAGACAATCTTAAGCAAAAAGAACAAAGCTGGAGGTATCATGCTTACCTGACTTAA 6514
Db 1390 AlaLysSerIleLeuSerGlnLysAsnLysAlaGlyGlyIleThrLeuProAspPheLys 1409
QY 6515 CTATACTAAGGCTACAGTAACCAAACTGCATGCTGCTACTGTTACCAAAACAGATATATA 6574
Db 1410 LeuTyrTyrLysAlaThrValThrLysThrThrTyrTrpTyrGlnHisArgAlaVal 1429
QY 6575 GACCAATGCAACAGAACAGAGACTCAGAAAATTTAC-ACTGCAATCTACATCCATCTGATC 6633
Db 1430 AspGlnTrpAsnArgThrGlyProSerGluIleMetProHisIleTyrAsnTyrLeuIle 1449
QY 6634 TTTGACAAACCTGCACAAAAACAGCAATGGAAGAGGATTCCTTATTTAATAATGGTGT 6693
Db 1450 PheAspLysProAspLysAsnLysGlnTrpGlyLysAspSerLeuPheAsnLysTrpCys 1469
QY 6694 TGGAAAAACCTGGCTAGCCATATGCCAAAGCTGAAAGCTGATCCCTTCTTACACCTTAT 6753
Db 1470 TrpGluAsnTrpLeuAlaIleCysArgLysLeuLysLeuAsnProPheLeuThrProTyr 1489
QY 6754 ACAAAAGTTAACTCAAGATGAATTAAGACTTAAATATAGACATAAAAACCAATAAAACC 6813
Db 1490 ThrLysIleAsnSerArgTrpIleLysAspLeuHisValArgProLysThrIleLysThr 1509

Db	176	sGluValTyrSerMetArgThrMetGlyArgArgGlnProGlyThrAlaMetAspLeuAs	196
QY	1719	TGAGGGAGGGGGCTGCTGCATTGCTGAGGCTTAGCGAGCGGGTTCACCTCAA	1778
Db	196	nAlaGlyGlySerIleCysAsnVal	204
QY	1779	AGTGTAAACAAGACTCTGGGAAGTTTGAATGGGGGCCACCGCAGCTCAGCAAGCCG	1838
Db	205	-----ProArgThrLeuLeuHisLeuThrGlyGluSerThrPheArgAspArgGlnAr	222
QY	1839	CTGTGCNAACTGCCCTCTCTAGATTCTCTCTTTTGGGAGGTCATCTCTGAAGAAGG	1998
Db	222	gValGly	231
QY	1899	CAGCAGCCCGCAGTCAGGACTTATAGATAAAACCCCATCTCCCTG---GGACAGAGA--	1953
Db	231	yTPrTyrIleLeuValLeuValValAspLysGluLysLeuGlyIleProGlyGlnLysG	251
QY	1954	-----ACCTGGGG-----AAGGGTGTCTGTGGTGCAGCT	1985
Db	251	nAsnSerSerLysIleLeuPheGlySerGluLysLeuLysGlnTrpValTyrValLys--	270
QY	1986	TCTCAGACTTAACATTCCTGCTGGAGGCTCTGAAGAGCAGCAGCGACTCCACGACA	2045
Db	271	-----As	271
QY	2046	GCATTTGAGCTCTGATAAGGACGACGCTGCTCTCAAGTGGCTCCCTGACCCCATGTA	2105
Db	271	nHisAlaLysGluSerIleThrGlnGluAlaMetGly	283
QY	2106	TCCTGACTGGAGACATCTCCCATTAGGGGCCAATAGACATTTACAGGAGACAGGCT	2165
Db	284	-----	284
QY	2166	CTGSAGTGGACCTCCAGCAACTCCAGCAGACCTCGCAGCAGCGCGCTGACTGTAGAA	2225
Db	284	etrpValGlyGln	288
QY	2226	GGAAAGTAAACAACAGAAAGGAATAGATPAAC-----ATTAACAA	2267
Db	289	-----GlnArgLysArgGlnProLeuGlyTyrGluGluArgLysLeuThrAsnAr	305
QY	2268	AAAGGACATCCATCAGACACCCATCTGAAGTCAACAACATCAAGACCAAGTA	2327
Db	305	glysAspIleHisIleLysAsnProSerValCysHisHisGlnArgProLysValas	325
QY	2328	TAAACCAAAAAGATGGAAAAAC-CAGTGCAGAAACACTGAAATTCCAAAAACAGAA	2386
Db	325	plysThrThrLysMetGlyLysLysGlnSerArgLysThrGlyAsnSerLysAsnGlnse	345
QY	2387	CTCCTCTCTCAACCAAGGATCACAACCTCCTCGCAGCAAGGAAACAAACCATGGA	2446
Db	345	rThrSerProThrLeuLysGluSerSerSerProAlaThrGluGlnSerTrpThrG	365
QY	2447	GAATGAGTTGAGGAATTGACAGAAGTAGGCTTCAGAAGTGGGTATACAAACCTCCT	2506
Db	365	uasnAspPheAspGluLeuArgGluGluGlyPheArgArg-----SerAspTyrSe	382
QY	2507	CGAGCTAAAGGACGATCTTCAACCAATGTCAGGAAGCTTAAGAACCCTTCAAAAAGGTT	2566
Db	382	rGluLeuGlnGluValGlnAsnAsnGlyLysGluValLysAsnPheGlyLysLysLe	402
QY	2567	AGATGAATTCCTAACTAGAATAATCAGTGTAGAGAAGAACATAATACCTGTAGAGCT	2626
Db	402	uasgGluTrpIleThrArgIleThrAsnAlaLysLysSerLeuLysAspLeuMetGluLe	422
QY	2627	GAACAACGCAAGCAACTTCATGAAGCATACAAAGCTCAATAGCCAAATCGATCA	2686
Db	422	uLysThrLysAlaArgGluLeuLysAspGluArgThrSerLeuSerSerGlnCysAsnG	442
QY	2687	AGCAGAAGAAGATATCAGTATTGAAGTCAAAATTAATAAAGAAGAGTCAAGACAA	2746
Db	442	nLeuGluGluArgValSerValMetGluAspGluMetAsnGluIleLysGlnGluGly	462
QY	2747	GATTACAGAAAAAGAGTGAAAGAAACAAACAAAGCCTCCAAGAATTTATGGACTATGT	2806
Db	462	sPheArgGluLysArgIleLysArgAsnGluGlnSerLeuGlnGluIleTrpAspTyrMe	482
QY	2807	GAAGAACCAATCTCATTTGATTTGTTCCCTCCCAAGATGATGGGAGAAATGAATCAA	2866
Db	482	tLysArgProAsnLeuArgLeuIleGlyValProGluSerAspGlyGluAsnGlyThrLy	502
QY	2867	GTTCGAAAACACTCTTCAGGTTATTCAGGAGAAATTTCCCAT-CTATCAGGGCAGGC	2925
Db	502	sLeuGluAsnThrLeuGlnAspIleIleGlnGluAsnPheProAsnLeuAlaArgGlnAl	522
QY	2926	CAACATTCAAATTCAGAGAAATATGGAGAACACCAATAGATATCTCTCCAGAGAACAAAT	2985
Db	522	asnIleGlnIleGlnGluIleGlnArgThrProGlnArgTyrSerLeuArgArgAlaTh	542
QY	2986	CCCAAGACACATATCTCAGATTCCCAAGGTTGAAATGAAGAAATAATTTAGAGGC	3045
Db	542	rProArgHisIleIleValArgPheThrLysValGluMetLysGluLysMetLeuArgVa	562
QY	3046	AGCCAGAGAGAAAGTTGGTTACCCACAAAGGGAAGCCCAATCAGACTAACACGGATCT	3105
Db	562	lAlaArgGluLysGlyArgValThrHisLysGlyLysProValArgLeuThrAlaAspLe	582
QY	3106	CCCGCAGAAACCTTACAAGCCAGAGAGAGTGGGGCCCAATATTCACATTTCTTAAAGA	3165
Db	582	uLeuAlaGluThrLeuGlnAlaArgGln	592
QY	3166	AAATAATTTTCAACCCAGAAATTCATATCCAGCCAACCAAGCTTCTTAAGTGAAGAGA	3225
Db	592	-----	592
QY	3226	AATAAATCCTCTACAGAGAAGCAATGCTGCACAGATTTTGTGCACCACGAGCCTGCCT	3285
Db	592	-----	592
QY	3286	TACAAGAGCTCTGAAGGAAGCACCACATGGAAGGAACAACACTGTTACCAGCCACTGCA	3345
Db	592	-----	592
QY	3346	AAACATCCCAAAATTGTAAGACCATTGATGCTATGAGAAAGTGATCAACTAACGGGC	3405
Db	592	-----	592
QY	3406	AAAAAACACGCTAGTGTCTAATATGCAGGATCAAAATTCACACATAATAATATAACCTT	3465
Db	592	-----	592
QY	3466	AAATGTAAATGGCTAAATTCCTCAATTAAGAGACACAGACTGGCAAAATTTGGATAAGAG	3525
Db	592	-----	592
QY	3526	TCAAGACCCTCAGTGTGCTGATTTCAGGAGGCCCATCTCACATGAAAGACACACATAG	3585
Db	593	-----AspThrHisar	596
QY	3586	GCTCAAAATAAAGGATGGAGGAAGATTACCAAGTAAATGGAAGAAAAAAGG	3645
Db	596	gLeuLysIleLysGlyTyrArgLysIleThrGlnAlaAsnGlyLysGln---LysLysAl	615
QY	3646	AGGGTGTCAATCTCTGCTCTGATTAACACAGACTTTAAACCAACAAAGATCAAAAGAGA	3705
Db	615	agLysAlaIleLeuValSerAspLysThrGluPheLysProThrLysIleLysArgas	635
QY	3706	CAAGAAGGCCATTACATATGCTTAAGCATCAATGGAACAAGACAGCTACTACTCT	3765
Db	635	pLysGluGlyHisTyrIleMetValLysGlySerIleGlnGlnGluLeuThrIleLe	655
QY	3766	AAATATACATGCACCAATACAGCAGCCACATTCATAAACAAAGTCTCTTAGAGACCT	3825
Db	655	uasnIleTyrAlaProAsnThrGlyAlaProArgPheThrLysGlnValLeuArgAspLe	675

```
QY 3826 ACAAGAGAGACTTTGACTCCCAACACAAATAATAGTGGAGTCTAAATAATAATAGACACTT 3885
Db 675 uGlnArgAspLeuAspSerAsnThrIleIleThrGly-----AspH 689
QY 3886 TAACACCCACATGCCAATATTAGGACGATCAATAGACAGAGAAATTAACAAGGATATCCA 3945
Db 689 easnThrProLeuSerThrLeuAspArgSerMetArgGlnLysValAsnLysAspIleG1 709
QY 3946 GGAGTTGAACCTGAGCTCGGACCAAGCGACCTAATAGATATCTACAGAACTCCGCCACCC 4005
Db 709 nGluLeuAsnSerAlaLeuHisGlnAlaAspLeuIleAspIleTyArgThrLeuHisP 729
QY 4006 CAAATACAGAAATATACACTCTCTCAGCATCATACATACACTATTTTAAATATGACCA 4065
Db 729 oLysSerThrGluAsnThrPhePheSerAlaProHisHisThrTySerLysIleAspHi 749
QY 4066 TGTAAATTTTAACTAAACACTCTCAGCAAAATGCAAAAGACAGAAATCCTTAACAAACAG 4125
Db 749 sIleValGlySerLysAlaLeuLeuSerLysCysLysArgLysGluIleIleThrAsnCy 769
QY 4126 TCTCTCAGACTACAGTGAATCTATTAGAACTCAGAATTAAGAAACTCACTCAAAATCA 4185
Db 769 sLeuSerAspHisSerAlaIleLysLeuGluLeuArgIleLysLysLeuProGlnAsnC 789
QY 4186 CAACTACATGGAACCTGACAACTGCTCTGCTGAATGACTACTGGGTAAATAACAAAT 4245
Db 789 sSerThrIleTrpLysLeuAsnAsnLeuLeuLeuAsnAspTyTrpValHisAsnGluMe 809
QY 4246 GAAGCAAAATAAAGATGTTCTTTGAACCAATGAGAAACAGACAAATGTACCAGAA 4305
Db 809 tLysAlaGluIleLysMetLeuPheGluThrAsnGluAsnLysAspThrThrTyrgLnas 829
QY 4306 TCTCTGGGCGATATTAAAGCAGTGTGTAGAGGAAATTTATAGCACTAGATGCCATCAA 4365
Db 829 nLeuTrpAspThrLeuLysAlaValCysArgGlyLysPheIleAlaLeuAsnAlaHisG1 849
QY 4366 GAGAAAGCAGAAATATCTAAATAGACACCTTAACTACATCACAATTAAGAAGACTAGAGAA 4425
Db 849 yArgLysGlnGluArgSerLysIleAspThrLeuThrSerGlnLysGluLeuAlaLys 869
QY 4426 GAAAGCAAAATAAAGCTAGCAGAGACAGAAATAACTAAGATCAGACGACAGA 4485
Db 869 sGlnGluGlnThrHisSerLysAlaSerArgGlnGluIleThrLysIleArgAlaG1 889
QY 4486 ACTGAAGGAGATAGACACAAAAGCCCTTCAATAAATCAATGAATCCAGGAGCTGGT 4545
Db 889 uLeuLysGluIleGluThrGlnLysThrLeuGlnLys-IleAsnGluSerArgSerTrpp 909
QY 4546 TTTTGTAAAGATCAGCAAAATAGACCCTAGACAGACTAATAAGAGAAAGAGAGAA 4605
Db 909 he-----Leu--ProLeuAlaArgLeuIleLysLysLysThrGluLys 922
QY 4606 GAATCAAAAGAGATGCAATATAAAATGATAAAGGGGATATACACCGATCCACAGAAAT 4665
Db 922 sasGlnIleAspAlaIleLysAsnAspLysGlyAspIleThrThrAsnProThrAspI1 942
QY 4666 ACAACTATATCAGAAATATTATAACACCTCTATGCAATTAACCTAGAAAATCTAGA 4725
Db 942 eGlnThrAlaIleArgGluTyTyTyLysHisLeuTyAlaAsnLysLeuGluAsnLeuG1 962
QY 4726 AGAAATGATAAATCTCTGGACACATATGTAGACCTGTATGACCTTGGGGACAGACAA 4785
Db 962 uGluMetAspLysLeuLeu----- 968
QY 4786 AAGGGGTGAATGCAGAAATAAAGACAAAGACAAAGATGATGTTTGGAAAGTAGGGTCC 4845
Db 968 ----- 968
QY 4846 AGGGGGCAACTTGCTCTAATGGACAAGGGCCCTGAGCTTTACACCACTCTGTATTTA 4905
Db 968 ----- 968
QY 4906 TTAGGCAAAAGAGATACGAGAGGGTGAGTTTGGAAAGAGAGGTGAGCTGTAGTCCAGA 4965
```

```
Db 968 ----- 968
QY 4966 GTAGCCCTGCAAGACTGCATTCCTCAAACAATAGCTCTAGATGTCCCCAGTAGATAACCT 5025
Db 968 ----- 968
QY 5026 CAAAGGAGCAGTGCAGGGAGTGATGGCCCTCAGCAAAACCTTCTAGGGCAGGCACAGAAG 5085
Db 968 ----- 968
QY 5086 TAAGTTTCCCACATCTGTATTTCAGATAAACACTTGTGCTGTTTGTATCAAGTAGACCTCC 5145
Db 968 ----- 968
QY 5146 AGTGAATGCTCAGTTGTCATGATCCCTTTTGGCCCTTTTGGCTCCCAAAACACATACAC 5205
Db 969 -----AsnThrTyTrh 972
QY 5206 CTTCTCAAGACTAAACACAGGAAGTCAAAATCCCTGAATATACCAAGTAAACAAAGTTCTAA 5265
Db 972 rIleProArgLeuAsnGlnGluGluValGluProArgAsnArgProIleThrGlySerG1 992
QY 5266 AATTGAACGAGTAATTGATAGCTTACCACCAACCAAAAAAGTCCAGGACGACGAGTATCAC 5325
Db 992 uIleGluAlaIleIleAsnArgLeuProThrLysLysCysProGlyProAspGlyPheTh 1012
QY 5326 AGCCAAATTTACCAGAGGTACAAAGAGAGAGTGTCTACTATTCTCTGAAACTATTCCA 5385
Db 1012 rAlaGluPheTyrgLnaArgTyrgLysGln-----ValProPheLeuLeuLysLeuPheG1 1030
QY 5386 AAAAATAGAA---AATGGGAATCCTCCCTAACTCATTTTACGAGCCAGCATCATCTCGA 5442
Db 1030 nSerThrGluLysGly-GlyIleLeuProLysSerPheTyrgLnaAlaSerIleIleLeu1 1050
QY 5443 TACCAAAACCTAGCAGTGACACACACAAAAAGAGGAAATTTTACGGCCCATATCCCTGATGA 5502
Db 1050 leProLysProGlyLysAspThrThrLysLysGluAsnPheArgProIleSerLeuMeta 1070
QY 5503 ACATTGATGTAAATCCTCAATAAAATCTGGCAAAACAAATCCAGCAGACATCAAAA 5562
Db 1070 snIleAspValLysIleLeuAsnLysIleLeuAlaLysGlnIleGlnHisIleLysL 1090
QY 5563 AGCTTATCTACCATGATCAAGTTGGCGTCATCCCTGGGATGCAAGCTGGTGTCAAAATAT 5622
Db 1090 ysLeuIleHisHisAspGlnValGlyPheIleProGlyMetGlnArgTrpPheAsnIleA 1110
QY 5623 GCAATCAATAATGTAGGCCCATCACATAAAGACAGAACCAATGACAAAACACATGATTA 5682
Db 1110 rgLysSerIleAsnValIleGlnHisIleAsnArgThrLysAspLysAsnArgMetIleI 1130
QY 5683 TCTCAATAGATGCAGAAAGGCCCTTGTCAAAATTCACACGCCCTTCATGCTAAAAATTC 5742
Db 1130 leSerIleAspAlaGluLysAlaPheAspLysIleGlnArgPheMetLeuLysThrL 1150
QY 5743 TCAGTAACTAGTATCGATGGAATGATCTCAAAATAAATAAGAGCTATTTATAC-AAAC 5801
Db 1150 euAsnLysLeuGlyIleAspGlyThrTyrgLysLysThrIleArgAlaIleTyrgLysLys 1170
QY 5802 CCAGCCCAATATCATCTGAATGGCAAAACCTGGAGAGCATTCCTTTGAGAACTGGCA 5861
Db 1170 roThrAlaAsnIleIleLeuAsnGlyGlnLysLeuGluAlaPheProLeuLysThrGlyT 1190
QY 5862 CAAGACAAGGATGCCCTCTCTCACCACCTCTTATTCAAGATACTATTGGAAGTCTTGGCCA 5921
Db 1190 hrArgGlnGlyCysProLeuSerProLeuLeuPheAsnMetValLeuGluValLeuAla 1210
QY 5922 GGGCAATCAGGCAATAGAAAGATAAAGGGTATTCAAAATAAGAGAGAGAGAGTATAT 5981
Db 1210 rgAlaIleSerGlnGluLysGluIleLysAspIleGlnLeuGlyLysGluGluValLysL 1230
QY 5982 TGTCTCTGTTGACAGATGACATGTTTGTATATTTAGAAAACCCCATCTCTCAGGCCAAA 6041
Db 6041
```

Db 1230 euSerLeuPhealaAspMetIleValTyLeuGluAsnProIleValSerAlaGlnA 1250
QY 6042 AACTCCTTAAAGCTGATAGCAAACTTCTCAGCAACAAATCAATGTCACAA 6101
Db 1250 snArgLeuLysLeuIleSerAsnPheArgLysValSerGlyTyrLysIleAsnValGlnL 1270
QY 6102 AATCACAAGCATCTTATACGCCAATAATAGACAACAGAGAGCCAAATCATAGTGAAC 6161
Db 1270 ysSerGlnAlaPheLeuTyrThrAsnAsnSerGlnThrGluSerGlnIleMetSerGluL 1290
QY 6162 TCTCATTCACAATTCCTCAAGAGAGATAAATACCTAGGATACAACTTACAGGGACA 6221
Db 1290 euProPheThrThrAlaSerLysArgIleLysTyrLeuGlyIleGlnLeuThrArgAspV 1310
QY 6222 CGTAGGAATCTTCAAGAGAGAACTACAAACCACTGATCAAGGAAATAAGAGAGACACAA 6281
Db 1310 aLysAspLeuPheLysGluAsnTyrLysProLeuLeuSerGluIleLysGluAspThra 1330
QY 6282 ACAATGGAAAACATTCATGCTCCTCACAGATA-----GTAAGAATCATGAAATGCC-A 6334
Db 1330 snLysCysLysAsnValProCysSerTrpValClyArgIleAsnIleMetLysMetAlaI 1350
QY 6335 TACTGCCAACTAAATATATACATTCAGTCTACCCCATCAAGTACCATTCGACTTCT 6394
Db 1350 leuProLysValIleTyrArgPheAsnAlaIleProIleLysLeuProMetThrPheP 1370
QY 6395 TCACAGAATGTGAAAAACAACATTTAAATTTTCATATGAACCAAAAAAGAGCCACACAGA 6454
Db 1370 heThrGluLeuGluLysThrLeuLysPheIleTrpAsnGln-LysArgAlaCysIle 1389
QY 6455 GCCACAGACANTTTAAGCAAAAAGAACAAAGCTGGAGGTATCATGCTACCTGACTTAAAA 6514
Db 1390 AlaLysSerIleLeuSerGlnLysAsnLysAlaGlyIleThrLeuProAspPheLys 1409
QY 6515 CTATACTATAAGCTTACACTACCAACAACTGGTACTGGTACTGTACCAACACAGATATATA 6574
Db 1410 LeuTyrTyrLysAlaThrValThrLysThrTrpTyrTrpTyrGlnHisArgAlaVal 1429
QY 6575 GACCAATGGAACAGACAGACCTCAGAAATTTAC-ACTGCAATCTACATCATCTGATC 6633
Db 1430 AspGlnTrpAsnArgThrGlyProSerGluIleMetProHisIleTyrAsnTyrLeuIle 1449
QY 6634 TTTGACAAACCTGCAAAAACAGCAATGGAAAAAGGATTCCTATTTAATAAATGGTGT 6693
Db 1450 PheAspLysProAspLysAsnLysGlnTrpGlyLysAspSerLeuPheAsnLysTrpCys 1469
QY 6694 TGGAAAACTGCTAGCCATATGCAGAAAGCTGAACCTGGATCCCTCTTACACCTTAT 6753
Db 1470 TrpGluAsnTrpLeuAlaIleCysArgLysLeuLysLeuAsnProPheLeuThrProTyr 1489
QY 6754 ACAAAGTTAACTCAAGATGAATTAAGACTTTAAATATAAGACATATAAACCAACCAACC 6813
Db 1490 ThrLysIleAsnSerArgTrpIleLysAspLeuHisValArgProLysThrIleLysThr 1509
QY 6814 CA-GAAGAAACCTAGGCAATACCATTCAGGATATGGACATGGGCAAGACTTCATGACT 6872
Db 1510 LeuGluGluAsnLeuGlyAsnThrIleGlnValIleGlyMetGlyLysAspPheMetSer 1529
QY 6873 AAACACCAAAAGCAATGGCAACAAACCAAAATAGACAAGTGGGATCTGATTAACATA 6932
Db 1530 LysThrProLysAlaMetAlaThrLysAlaLysIleAspLysTrpAspLeuIleLysLeu 1549
QY 6933 TAGAGCTTCTGCACAGCAAAAAAAGCTGTCATCAGAGTGAACAGCAACCTACAGAAATG 6992
Db 1550 LysSerPheCysThrAla-LysGluSerThrIleArgValAsnArgGlnProThrGluTyr 1569
QY 6993 GGAGAAAAATTTTGCAATCTTCGATCTGCACAAAGGCTAATATCCAGAGATCTACGAAGA 7052
Db 1569 pcuLysIlePheAlaThrTyrSerSerAspLysGlyGlnIleSerArgIleTyrAsnGln 1589
QY 7053 ACTTAAACAATTTTACAGAAAAA-----AACACCCCTCAAAATATGCGCAAGCATAT 7108
Db 1589 uLeuLysGlnIleTyrLysLysThrAsnAsnProIleLysLysTrpAlaLysAspAr 1609

QY 7109 GAGCAGACACTTCTCAAAAGAGACATTATGCAGCCAAACAAACATATGAAAAAACCTC 7168
Db 1609 gAsnArgHisPheSerGluGluAspIleTyrAlaAlaLysLysHisMetLysLysCysSe 1629
QY 7169 ATCATCATTTGGTCGTAGAGAAATGAAACAAAACACAGTACATACCATCTCATGCT 7228
Db 1629 rProSerLeuAlaIleArgGluValGlnIleLysThrMetArgTyrHisLeuThrPr 1649
QY 7229 AGTTAGAATGGTGATCACTATAAAGTTCAGGAACACACAA 7268
Db 1649 oValArgMetAlaIleLysLysSerGlyThrAsnArg 1662

RESULT 14
ABG08641
ID ABG08641 standard; Protein; 1691 AA.
XX
AC ABG08641;
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #8632.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS72828.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity -
XX
PS Claim 20; SEQ ID No 39000; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX

```
SQ Sequence 1691 AA;
Alignment Scores:
Pred. No.: 0 Length: 1691
Score: 4893.50 Matches: 1124
Percent Similarity: 60.58% Conservative: 147
Best Local Similarity: 53.57% Mismatches: 313
Query Match: 9.35% Indels: 522
DB: 22 Gaps: 32

US-10-083-853B-2 (1-29921) x ABG08641 (1-1691)
QY 1210 CATTCCTCCCACTGAGGTACCCAGTTCATCTC----- 1239
Db 8 HisProAlaGluTrpProArgHisValLeuLeuLeuMetAlaAspSerLysAsnGlu 27
QY 1240 -----ACTGGGACTGGTTAGACATTGGGTGCGAGCCCGACGAGAGGTGAGCTGAA--- 1287
Db 28 LeuGlnAsnMetAsnThrGlyGlnThrCysGlyThrArgGlnThrLeuLysSerLysLeu 47
QY 1288 GCAGGGTGGGTGTCCTCCCTCAGCCGCGAAGTGAAGGGGTGGGGGATCTCTCTCCGCCCA 1347
Db 48 GlnGlyTrpAsnLysPro-----GlyIleGlnSerAspPro 59
QY 1348 CCCAAGGAGCCATGAGAGACTGTACCAGAGGAGATGGTGCATCTAGTCCAGATACTG 1407
Db 60 GlySerGlnGlyHisProArg-CysLeuLys----- 69
QY 1408 CACTTTTCCCATAGTCTTTGCACTGGCAGACAGGAGATTTCCCGCAGTCCCTATGCCA 1467
Db 70 -ThrPheProLeuValAlaGlnSerSer-----ThrGlnGluArgAsnH1 84
QY 1468 C-----CAGGGCCCTGGGTTTCAAGCACAAAACACTGGCGGCCATT----- 1507
Db 84 svalmetGluLysAlaValLeuLysSerGlnAsnTrpAlaIleIleGluGluPheIleTh 104
QY 1508 -----TGGACAGACCCGAGCTAGCCGCGAGCAGTATTATTTTCATACCCAGTGG 1557
Db 104 rLysIleGlyTrpPheProAspTyrSerGlyLysValLysTyrLeuSerProLysSe 124
QY 1558 CGCTTGAATCCAGCAAGACAGAACCATTCATCCAGGGATCCAAAGTGGTCTGGCTCAG 1617
Db 124 rSer-----CysThrValLeuGlnValThrHisPheSe 136
QY 1618 T---GGGTCCACCCCATGAGCCAGCTA---GCTAAGATCCAC----- 1657
Db 136 rLeuPhePheHisProAsnProAspGluLeuValSerLysIleLysValTrpSerLysH1 156
QY 1658 -----TGGCTTGAATTCCTCTGCGCAGCAGCAGCAGTC 1689
Db 156 sArgLeuTyrGlnAsnAsnSerGluAlaPheLeuGluValGlnIleProGluProLysCy 176
QY 1690 TGAGATTGA-----CCTGGG---ATGCTTGAGCTGG 1718
Db 176 sGluValTyrSerMetArgThrMetGlyArgGlnProGlyThrAlaMetAspLeuAs 196
QY 1719 TGAGGGAGGGCGCTGTGCCATTGCTGAGGCTTGAGTAGCGGCGGCTTTTACCCCTCAA 1778
Db 196 nAlaGlyCysIleCysAsnVal----- 204
QY 1779 AGTGTAACAAGACTACTGGGAAGTTGAATGGGGCGCCACCCAGCTCAGCAAGCGCG 1838
Db 205 -----ProArgThrLeuLeuHisLeuThrGlyGluSerThrPheArgAspArgGlnAr 222
QY 1839 CTGTGGCAAACTGCTCTCTAGATTCCTCTCTTTTGGCAGGTCTCTCTCAAGAAAGG 1898
Db 222 gValGly-----AlaTyrLeuAspIleArgG1 231
QY 1899 CAGCAGGCCAGTCAGGAGCTTATAGATAAAACCCCATCTCCCTG---GCACAGAGA--- 1953
Db 231 yTrpTyrIleLeuValLeuValAspLysGluLysLeuGlyIleProGlyGlnLysG1 251
QY 1954 -----ACCTGGGGG-----AAGGGTGGCTGTGGGTGCAGCT 1985

251 nAsnSerSerLysIleLeuPheGlySerGluLysLeuLysGlnTrpValLys-- 270
1986 TCTCAGACTTAACATTCCTCGCTGGAGGCTCTCAAGAGAGACGCGGATCTCCACGACACA 2045
271 -----As 271
2046 GCATTGAGCTCTGATAAGGACAGCGTGCCTCTCAAGTGGGTCTCCTGACCCCATGTA 2105
271 nHisAlaLysGluSerIleThrGlnGluAlaMetGly----- 283
2106 TCCTGACTGGGAGACATCTCCATTAGGGGCCAATAGACATTTTCATACAGGAGACAGGT 2165
284 ----- 284
2166 CTGGAGTGGACCTCCAGCAAACTCCAGCAGACCTGCAGAGAGCGGCTGACTGTAGAA 2225
284 eTrpValGlyGln----- 288
2226 GGAAGAGTAACAACACAGAAAGGAATAGTATCAAC-----ATTAACA 2267
289 -----GlnArgLysArgGlnProLeuGlyTyrGluGluArgLysLeuThrAsnAr 305
2268 AAAGGACATCCACTCAGAGACCCCATCTGAAGGTCAACACATCAAGACCAAGGTAAA 2327
305 gLysAspIleHisLysAsnProSerValCysHisHisGlnArgProLysValas 325
2328 TAAACCAAAAAGATGGAAAAAC-CAGTGCAGAAACACTGAAATTTCCAAAAACAGAA 2386
325 pLysThrThrLysMetGlyLysLysGlnSerArgLysThrGlyAsnSerLysAsnGlnSe 345
2387 CTCCTCTTCTCAACAAAGGATCAACTCTCTCCAGCAAGGAGGAAACAAACAGATGGA 2446
345 rThrSerProThrLeuLysGluSerSerSerProAlaThrGluGlnSerTrpThrG1 365
2447 GAATGAGTTTCAGAAATTCAGAGAGTAGCTTCAGAGGTGGGTGTAATAACAACCTCCTC 2506
365 uAsnAspPheAspGluLeuArgGluGlyPheArgArg-----SerAspTyrSe 382
2507 CGACTTAAAGGAGCATGTTCTTAACCCCAATGCAAGAAAGCTTAAGAACCTTGAANAAGGTT 2566
382 rGluLeuGlnGluValGlnAsnAsnGlyLysGluValLysAsnPheGlyLysLysLe 402
2567 AGATGAATTCCTACTAGATAATCATCTGTAGAGAGAAACATAAATGACCTGTAGGAGCT 2626
402 uAspGluTrpIleThrArgIleThrAsnAlaLysLysSerLeuLysAspLeuMetGluLe 422
2627 GAAAAACGACAGACAACTTTCATGAAGCATACACAAAGCTTCAATAGCCAAATTCGATCA 2686
422 uLysThrLysAlaArgGluLeuCysAspGluArgThrSerLeuSerSerGlnCysAsnG1 442
2687 AGCAGAAAGAGGATATCAGTGTGAGATCAATTAATAAAGAAAGTGAGAGACAA 2746
442 nLeuGluGluArgValSerValMetGluAspGluMetAsnGluIleLysGlnGluLys 462
2747 GATTACAGAAAAGAGTGAAGAAACAAACAAAGCTCCCAAGATTTATGGGACTATGT 2806
462 sPheArgGluLysArgIleLysArgAsnGluGlnSerLeuGlnGluIleTrpAspTyrMe 482
2807 GAAAGACCAATCTACATTTGATTGTTGTCCTCCCAAGTGCATGGGAGATGGAATCAAA 2866
482 tLysArgProAsnLeuArgLeuIleGlyValProGluSerAspGlyGluAsnGlyThrLy 502
2867 GTTGAAAAACACTCTTCAGGTATTATCCAGGAGAAATTTCCCAT-CTATCAGGCGCAGGC 2925
502 sLeuGluAsnThrLeuGlnAspIleIleGlnGluAsnPheProAsnLeuAlaArgGlnAl 522
2926 CAACATTCAAATTCAGGAAATATGGAGAACACACCACTAAGATACCTCTCGAGAGAAACAAT 2985
522 aAsnIleGlnIleGlnGluIleGlnArgThrProGlnArgTyrSerLeuArgArgAlaTh 542
2986 CCCAAGACACATATCTTCAGATTCACCAAGGTGAATGAGGAAAAAATGTTAAGGC 3045
```

Db	542	rProArgHisIleIleValArgPheThrLysValGluMetLysGluLysMetLeuArgVa	562
QY	3046	AGCCAGAGAAAGGTGGTTACCCACAAAGGGAAGCCAAATCAGACTAACAGCGGATCT	3105
Db	562	lAlaArgGluLysGlyArgValThrHisLysGlyLysProValArgLeuThrAlaAspLe	582
QY	3106	CCCGCAGAAACCCCTACAGCCAGAGAGAGTGAAGGCCCAATATTCCACATCTTTAAAGA	3165
Db	582	uLeuAlaGluThrLeuGlnAlaArgGln	592
QY	3166	AAATAATTTTCAACCCAGAAATTCATATCCAGCCAAACCAAGCTTCCTAAGTGAAGGAGA	3225
Db	592		592
QY	3226	AATAAAATCCTCTACAGAGAAGCAAAATGCTGACAGATTTTGTCCACCACGCGCTGCT	3285
Db	592		592
QY	3286	TACAAGAGCTCCTGAAGGAAGCACCACCAATGGAAGGAACAACACTGGTACCAGCCACTGCA	3345
Db	592		592
QY	3346	AAACATCCCAAAATGCTAAGACCATTGATGCTATGAAGAAAGTGCGATCACTAACGGGC	3405
Db	592		592
QY	3406	AAAAATACACAGCTAGTGTCTAATATGCCAGGATCAAAATTCACACATAATAATTAACTT	3465
Db	592		592
QY	3466	AAATGTAATGGCTAAATTCOCCCAATTAAGACACAGACTGGCAAAATTTGGATAAAGAG	3525
Db	592		592
QY	3526	TCAAGACCCTCAGTGTGTGTATTTCAGAGGCCCATCTCACATGAAGAAGACACACATAG	3585
Db	593		596
QY	3586	GCTCAAAATAAAGGGATGAGGAAGATTTACCAAGTAATGGAACAAAAAAGGCG	3645
Db	596	gLeuLysIleLysGlyTrpArgLysIleTyrGlnAlaAsnGlyLysGln---LysLysAl	615
QY	3646	AGGGTTGCAATCCTAGTCTCTGATAAAGACAGACTTTAAACCAACAAGATCAAAAGAGA	3705
Db	615	agLyValAlaIleLeuValSerAspLysThrGluPheLysProThrLysIleLysArgAs	635
QY	3706	CAAGAAGGCCATTACATAATGGTAAAGCATCAATGGAACAGAGAGCTAACTATCCT	3765
Db	635	plysGluGlyHisTyrIleMetValLysGlySerIleGlnGlnGluLeuThrIleLe	655
QY	3766	AAATATACATGCCCAATACAGGACGCCAGATTCATAAAGCAAGTTCTTAGAGACT	3825
Db	655	uAsnIleTyrAlaProAsnThrGlyAlaProArgPheThrLysGlnValLeuArgAspLe	675
QY	3826	ACAAGAGACTTTGACTCCCAACATAATAGTGGGAGTCTAAATAATAATAGACACTT	3885
Db	675	uGlnArgaspLeuAspSerAsnThrIleIleThrGly-----AspPh	689
QY	3886	TACACCCCACTGCCAATATTAGGCAGATCAATGACAGACAGAAATTAACAAGGATATCCA	3945
Db	689	eAsnThrProLeuSerThrLeuAspArgSerMetArgGlnLysValAsnLysAspIleG	709
QY	3946	GGAGTTGAAGCTGAGCTCGACCAAGCGACCTAATAGATATCTACAGAACTCCCAACC	4005
Db	709	nGluLeuAsnSerAlaLeuHisGlnAlaAspLeuIleAspIleTyrArgThrLeuHisPr	729
QY	4006	CAAAATCAACAGAAATATACACTCTCTCAGCATCATCATTAACCTATTTTAAATAATGACCA	4065
Db	729	olysSerThrGluAsnThrPhePheSerAlaProHisHisThrTyrlSerLysIleAspHi	749
QY	4066	TGTAATTTTAAATAAACACTCTCTCAGCAAAATGCAAAAGACAGAAATCCTTAAACAACAG	4125
Db	749	sIleValGlySerLysAlaLeuLeuSerLysCysLysArgLysGluIleIleThrAsnC	769

DT 13-FEB-2002 (first entry)
 XX Novel human diagnostic protein #10793.
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.
 KW Homo sapiens.
 XX W0200175067-A2.
 XX 11-OCT-2001.
 XX 30-MAR-2001; 2001WO-US08631.
 XX 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX (HYSE-) HYSEQ INC.
 PA Drmanac RT, Liu C, Tang YT;
 PI WPI: 2001-639362/73.
 DR N-PSDB; AAS74989.
 XX New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity
 XX Claim 20; SEQ ID No 41161; 103pp; English.
 XX The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 1691 AA;
 SQ
 Alignment Scores:
 Pred. No.: 0 Length: 1691
 Score: 4893.50 Matches: 1124
 Percent Similarity: 60.58% Conservative: 147
 Best Local Similarity: 53.57% Mismatches: 313
 Query Match: 9.35% Indels: 522
 DB: 22 Gaps: 32
 US-10-083-853B-2 (1-29921) x ABG10802 (1-1691)
 QY 1210 CATTCCCACTGAGGTACCGATTATCTC----- 1239
 DB 8 HisProAlaGluTrpProArgHisValLeuLeuMetAlaAspSerLysAsnGlu 27
 QY 1240 -----ACTGGGACTGTTAGACATTGGGTGCGCCACCGAAGGTGACGTGAA--- 1287
 DB 28 LeuGlnAsnMetAsnThrGlyGlnThrCysGlyThrArgGlnThrLeuLysSerLysLeu 47

QY 1288 GCAGGGTGGGTGTCCCTCAGCCGGAAGTGCAGGGGTGGGATCTCTCTCCCCCA 1347
 DB 48 GlnGlyTrpAsnLysPro-----GlyLeuGlnSerAspPro 59
 QY 1348 GCCAAGGGAAGCATGAGACTGTACAGAGGAATGTGCTACTCTAGTCCAGATACTG 1407
 DB 60 GlySerGlnGlyHisProArg-CysLeuLys----- 69
 QY 1408 CACTTTTCCCATAGTCTTTTCAACTGGCAGACCAGAGATTTCCTCCAGTGCCTATGCCA 1467
 DB 70 -ThrPheProLeuValAlaGlnSerSer-----ThrGlnGluArgAsnH1 84
 QY 1468 C-----CAGGGCCCTGGGTTTCAAGACACAAAACCTGGGGCCCATT----- 1507
 DB 84 sValMetGluLysAlaValLeuLysSerGlnAsnTrpAlaIleGluGluPheIleTh 104
 QY 1508 -----TGGACAGACACCGAGCTAGCCGAGCAGGTATTATTTTCATACCCAGTGG 1557
 DB 104 rLysIleGlyTrpPheProAspTyrSerGlyLysValLysTyrLeuSerSerProLysSe 124
 QY 1558 CGCTCGAATGCCAGCAGACAGAACCATTCACCTCCAGGATCCAAAGTGTCTGGCTCAG 1617
 DB 124 rSer-----CysThrThrValLeuGlnValThrHisPheSe 136
 QY 1618 T---GGGTCCCAACCCCATGGAGCCAGCTA---GCTAAGATCCAC----- 1657
 DB 136 rLeuPhePheHisProAsnProAspGluLeuValSerLysIleLysValTrpSerLysH1 156
 QY 1658 -----TGGCTTGAATTCCTCTCCGCCAGCAGCAGCAGTC 1689
 DB 156 sArgLeuTyrGlnAsnAsnSerGluAlaPheLeuGluValGlnIleProGluProLysCy 176
 QY 1690 TGAGATTGA-----CCTGGG---ATGCTCAGCTGG 1718
 DB 176 sGluValTyrSerMetArgThrMetGlyArgArgGlnProGlyThrAlaMetAspLeuAs 196
 QY 1719 TGAGGGAGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1778
 DB 196 nAlaGlyGlySerIleCysAsnVal----- 204
 QY 1779 AGTGTAAACAAAGCTACTGGAAGTTGAATGGGGCGCCACCGCAGCTCAGCAAGCCG 1838
 DB 205 -----ProArgThrLeuLeuHisLeuThrGlyGlySerThrPheArgAspArgGlnAr 222
 QY 1839 CTGTGCAAAACTGCCTCTCTAGATTCTCTCTTTTGGGCGAGGTCATCTCTGAAAGAAAG 1898
 DB 222 gValGly-----AlaTyrLeuAspIleArgG1 231
 QY 1899 CAGCAGCCCACTGAGGACTTATATATAAACCCTCCTCCTG---GGACAGAGA-- 1953
 DB 231 yTrpTyrIleLeuValLeuValValAspLysGluLysLeuGlyIleProGlyGlnLysG1 251
 QY 1954 -----ACCTGGGG---AAGGGGTGCTGTGGGTGCGAGCT 1985
 DB 251 nAsnSerSerLysIleLeuPheGlySerGluLysLeuLysGlnTrpValTyrValLys-- 270
 QY 1986 TCTCAGACTTAAACATTCCTGCTGAGGCTCTGGAAGAGCAGCGGATCTCCAGCACA 2045
 DB 271 -----AS 271
 QY 2046 GCATTTGAGCTCTGATAAGGGACAGCTGCTCTCCTCAAGTGGTCCCTGACCCCATGTA 2105
 DB 271 nHisAlaLysGluSerIleThrGlnGluAlaMetGly----- 283
 QY 2106 TCCTGACTGGGAGACATCTCCCATTTAGGGCCCAATAGACATTTTCATACAGGACAGGGT 2165
 DB 284 ----- 284
 QY 2166 CTGGAGCTGGACCTCCAGCAAACTCCAGCAGACCTGCAGCAGAGCGGCTGACTGTGTAGAA 2225
 DB 284 eTrpValGlyGln----- 288

Db 849 yArgLysGlnArgSerLysIleAspThrLeuThrSerGlnLeuLysGluLeuAlaLys 869
Qy 4426 GAAAGACAAACAAATCAAAAGCTAGCAGAGACAGAAATTAACTAAGATCAGAGCAGA 4485
Db 869 sGlnGluThrHisSerLysAlaSerArgArgGlnGluThrLysIleArgAlaG 889
Qy 4486 ACTGAGGAGATGAGACACAAAGCCCTTCAAAATAAATCAATGAATCCAGGAGCTGGT 4545
Db 889 uLeuLysGluLeuThrGlnLysThrLeuGlnLys-IleAsnGluSerArgSerTrp 909
Qy 4546 TTTTGAAGATGACGACAAATAGACCACCTAGACAGACTAATAAGAGAGAGAGAGAA 4605
Db 909 he-----Leu--ProLeuAlaArgLeuLysLysThrGluLys 922
Qy 4606 GAATCAAGAGATGCAATAAAATGATAAGGGGATATCACACCGCATCCACAGAAAT 4665
Db 922 sAsnGlnIleAspAlaIleLysAsnAspLysGlyAspIleThrAsnProThrAsp 942
Qy 4666 ACAAACTATTATCAGAGAATATTATAACACCTCTATGCAAAATAAACTAGAAAATCTAGA 4725
Db 942 eGlnThrAlaIleArgGluTyrLysHisLeuTyrAlaAsnLysLeuGluAsnLeuG 962
Qy 4726 AGAAATGGAATAATCTCTGACACATATGTAGCCTGTATGGACCTTGGGGGACAGACAA 4785
Db 962 uGluMetAspLysLeuLeu 968
Qy 4786 AAGGGGTGAATGCAGAAATAAAAGACAAAGAGAGTATGTTTGAAGTAGGGGTC 4845
Db 968 ----- 968
Qy 4846 AGGGGCACTTGCCTTAATGACAGAGGCCCTGAGCTTTACACCACCTCTGTATTTA 4905
Db 968 ----- 968
Qy 4906 TTAGGCAAAAGATAGCAGAGGGTGAGTTGGAAGAGAGGTGAGCTGTAGGTCAGA 4965
Db 968 ----- 968
Qy 4966 GTAGGCTGCAAGACTGCATTCTCTCAACAATAGGCTGTAGATGTCCTCAGTAGATAACCT 5025
Db 968 ----- 968
Qy 5026 CAAGGACCACTGCCAGGAGTGTATGCGCCCTCAGCAACCTTCTAGGGCAGGCACAGAG 5085
Db 968 ----- 968
Qy 5086 TAAGTTGCCACATCTGTATTCACGATAAACAGTTGCTGTTGTATCAAGTAGCCTCC 5145
Db 968 ----- 968
Qy 5146 AGTGAATGCTGAGTTGGTCATGATCCCTTTGGCCCTTTTGGCTCCCAAAACACATACAC 5205
Db 969 -----AsnThrTyrTh 972
Qy 5206 CCTCTCAAGACTAAACAGAGAGAGTCAAAATCCCTGAATATACAGTAACAGTCTAA 5265
Db 972 rIleProArgLeuAsnGlnGluValGluProArgAsnArgProIleThrGlySerG 992
Qy 5266 AATTGAGCAGTAATTGATAGCCTACCAACCAAAAAGTCCAGGACGAGGATTCAC 5325
Db 992 uIleGluAlaIleAsnArgLeuProThrLysLysCysProGlyProAspGlyPheTh 1012
Qy 5326 AGCCAATTTACAGAGGTACAAAGAGAGCTGGTACTATTCTCTCTGAACTATTCCA 5385
Db 1012 rAlaGluPheTyrGlnArgTyrLysGln-----ValProPheLeuLysLeuPheG 1030
Qy 5386 AAAATAGAA---AATGGGAATCCCTCAACTCATTTTACGAGGCGCAGCATCCTGA 5442
Db 1030 nSerThrGluLysGly-GlyIleLeuProLysSerPheTyrGluAlaSerIleIleu 1050
Qy 5443 TACCAAACTGACAGTGACACAAAGAGAGAAATTTACGGGCCATATCCCTGATGA 5502
Db 1050 leProLysProGlyLysAspThrThrLysLysGluAsnPheArgProIleSerLeuMeta 1070

Qy 5503 ACATTGATGTGAAATCCTCAATAAATACTGCAAAACCAATCCAGCAGCACATCAAAA 5562
Db 1070 snIleAspValLysIleLeuAsnLysIleLeuAlaLysGlnIleGlnHisIleLysL 1090
Qy 5563 AGTTTATCTACCATGATCAAGTTGGCGTCATCCTCGGATGCAAGGCTGGTTCAAAATAT 5622
Db 1090 ysLeuIleHisAspGlnValGlyPheIleProGlyMetGlnArgTrpPheAsnIleA 1110
Qy 5623 GCAAACTAATAATGTAGGCCATCACATAAACAGAACCAATGACAAAAACCATGATTA 5682
Db 1110 rGlySerIleAsnValIleGlnHisIleAsnArgThrLysAspLysAsnArgMetIle 1130
Qy 5683 TCTCAATAGATGCAGAAAAGCCCTTGTCAAAATTCAAACAGCCCTTCATCTCAAAAATTC 5742
Db 1130 leSerIleAspAlaGluLysAlaPheAspLysIleGlnArgPheMetLeuLysThrL 1150
Qy 5743 TCAGTAACTAGTATCGATGGAATGTATCTCAAAATAATAAGAGCTATTTATAC-AAAC 5801
Db 1150 euAsnLysLeuGlyIleAspGlyThrTyrLeuLysThrIleArgAlaIleTyrAspLys 1170
Qy 5802 CCACAGCCATATCATCTCAATGGGCAAAACTGGAGCAATCCCTTTGAGAACGCGCA 5861
Db 1170 roThrAlaAsnIleIleLeuAsnGlyGlnLysLeuGluAlaPheProLeuLysThrGly 1190
Qy 5862 CAAGCAAGGATGCCCTCTCTCACCACTCTCTATTCAAGATACTATTGGAAAGTTCTGGCCA 5921
Db 1190 hrArgGlnGlyCysProLeuSerProLeuLeuPheAsnMetValLeuGluValLeuAla 1210
Qy 5922 GGCAATCAGGCAATAGAAAGTAAGGGTATTCAAATAGAAAGAGAGAGCAAGTCATAT 5981
Db 1210 rAlaIleSerGlnGluLysGluIleLysAspIleGlnLeuGlyLysGluGluValLysL 1230
Qy 5982 TGTCTCTGTTGTCAGATGACATGTTGTATATTAGAAAACCCCTCTCTCAGGCGCAA 6041
Db 1230 euSerLeuPheAlaAspAspMetIleValTyrLeuGluAsnProIleValSerAlaGln 1250
Qy 6042 AACTCCTTAAGCTGATAAGCAACTCTCAGCAAGCTCTCAGCACACAAAATCAATGTGCAA 6101
Db 1250 snArgLeuLysLeuIleSerAsnPheArgLysValSerGlyTyrLysIleAsnValGlnL 1270
Qy 6102 AATCACAAGCATCTTATACCCCAATTAATAGACAACAGAGAGCCCAATCATGAGTAAC 6161
Db 1270 ysSerGlnAlaPheLeuTyrThrAsnAsnSerGlnThrGluSerGlnIleMetSerGluL 1290
Qy 6162 TCTCATTCACAATTCGTACAAAGACAGATAAATACCTAGCAATACACTTACAGGAGCA 6221
Db 1290 euProPheThrAlaSerLysArgIleLysTyrLeuGlyIleGlnLeuThrArgAspV 1310
Qy 6222 CGTAGGAACCTCTCAAGGAGAACTACAACCACTGATCAAGGAAATAAGAGAGAGACACA 6281
Db 1310 alLysAspLeuPheLysGluAsnTyrLysProLeuLeuSerGluIleLysGluAspThra 1330
Qy 6282 ACAATGGAAAAACATTCCTCATGCTCACAGATA-----GTAGAATCATGAAAAATGCC-A 6334
Db 1330 snLysCysLysAsnValProCysSerTrpValGlyArgIleAsnIleMetLysMetAlaI 1350
Qy 6335 TACTGCCCAAAGTAAATATTATAGATTTCAGTCTACCCCTCCATCAAGCTACCATTCACTTCT 6394
Db 1350 leLeuProLysValIleTyrArgPheAsnAlaIleProIleLysLeuProMetThrPheP 1370
Qy 6395 TCACAGAAATGGAAAAACAACCTTTAAATTTTCATATGGAACCAAAAAAGAGAGCCACAGA 6454
Db 1370 heThrGluLeuGluLysThrThrLeuLysPheIleThrAsnGln-LysArgAlaCysIle 1389
Qy 6455 GCGAAGCAATCTTAAGCAAAAAAGAACAGCTGGAGGTATCATGCTACTGCTACTGACTTAA 6514
Db 1390 AlaLysSerIleLeuSerGlnLysAsnLysAlaGlyIleThrLeuProAspPheLys 1409
Qy 6515 CTATACATTAAGGCTACAGTAACCAAACTGGTACTGTTGTTACCAAAACAGATATATA 6574
Db 1410 LeuTyrTyrLysAlaThrValThrLysThrThrTrpTyrTrpThrGlnHisArgAlaVal 1429

QY 6575 GACCAATGGAACAGACAGACACCTCAGAAATTTAC- ACTGCAATCTACATCCATCTGATC 6633
|||||
Db 1430 AspGlnTrpAspArgThrGlyProSerGluIleMetProHisIleTyrAsnTyrLeuIle 1449
|||||
QY 6634 TTTGACAAACCTGCACAAACAGCAAGCAATGGAAAAGGATTCCTTATTTAATAAATGGTGT 6693
|||||
Db 1450 PheAspLysProAspLysAsnLysGlnTrpGlyLysAspSerLeuPheAsnLysTrpCys 1469
|||||
QY 6694 TGGAAAACTGGCTAGCCATTCAGCAAGAGCTGGAACCTGGATCCCTTCTTACACCTTAT 6753
|||||
Db 1470 TrpGluAsnTrpLeuAlaIleCysArgLysLysLeuAsnProPheLeuThrProTyr 1489
|||||
QY 6754 ACAAAGTTAACTCAAGATGAATTAAGACCTTAAATATAAGACATAAAACCATAAAACC 6813
|||||
Db 1490 ThrLysIleAsnSerArgTrpIleLysAspLeuHisValArgProLysThrIleLysThr 1509
|||||
QY 6814 CA-GAAGAAACCTAGGCAATACCATTCAGGATATGGACATGGGCAAGACTTCATGACT 6872
|||||
Db 1510 LeuGluGluAsnLeuGlyAsnThrIleGlnValIleGlyMetGlyLysAspPheMetSer 1529
|||||
QY 6873 AAACACCAAAAGCAATGGCAACAAAGCCAAATAGACAGTGGGATCTGATTAAACTA 6932
|||||
Db 1530 LysThrProLysAlaMetAlaThrLysAlaLysIleAspLysTrpAspLeuIleLysLeu 1549
|||||
QY 6933 TAGACCTTCTGCACAGCAAAACAACTGTCTATCAGAGTGACACAAACCTACAGAATG 6992
|||||
Db 1550 LysSerPheCysThrAla-LysGluSerThrIleArgValAsnArgGlnProThrGluTr 1569
|||||
QY 6993 GGAGAAATTTTGGCAATCTCGATCTGACAAAGGCTAATATCCAGAGATCTACGAAGA 7052
|||||
Db 1569 pGluLysIlePheAlaThrTyrSerAspLysGlyGlnIleSerArgIleTyrAsnG1 1589
|||||
QY 7053 ACTTAACAAATTTACAGAAAAA---AACAAACCCGTCAAAATATGGCAAGGATAT 7108
|||||
Db 1589 uLeuLysGlnIleTyrLysLysLysThrAsnAsnProIleLysLysTrpAlaLysAspAr 1609
|||||
QY 7109 GAGCAGACACTTCTCAAAGAGACATTTATGACGCCCAACAAACATATGAAAAAACCTC 7168
|||||
Db 1609 gAsnArgHisPheSerGluGluAspIleTyrAlaAlaLysLysHisMetLysLysCysSe 1629
|||||
QY 7169 ATCATCATTTGGTTCGTACAGAAATGCAAAACAAACACAGTGCATACCATCTCATGCT 7228
|||||
Db 1629 rProSerLeuAlaIleArgGluValGlnIleLysThrThrMetArgTyrHisLeuThrPr 1649
|||||
QY 7229 AGTTAGAATGGTGATCACTAAAAAGTCAGGAAACAAACAAA 7268
|||||
Db 1649 oValArgMetAlaIleIleLysLysSerGlyThrAsnArg 1662
|||||

Search completed: July 4, 2003, 20:07:14
Job time : 1229.5 secs